(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 June 2004 (03.06.2004)

PCT

(10) International Publication Number
WO 2004/046189 A2

(51) International Patent Classification[7]: C07K 16/00

(21) International Application Number:
PCT/GB2003/004964

(22) International Filing Date:
14 November 2003 (14.11.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0226731.8 15 November 2002 (15.11.2002) GB

(71) Applicant *(for all designated States except US)*: **MEDICAL RESEARCH COUNCIL** [GB/GB]; 20 Park Crescent, London W1B 1AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **RABBITTS, Terence, Howard** [GB/GB]; MRC Laboratory of Molecular Biology, Division Of Protein and Nucleic Acid Chemistry, Hills Road, Cambridge CB2 2QH (GB). **CHUNG, Grace** [CN/CN]; Department of Chemical Pathology, The Chinese University of Hong Kong, The Prince of Wales Hospital, Shatin, New Territories, Hong Kong (CN). **TANAKA, Tomoyuki** [JP/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB). **LOBATO-CABALLERO, Maria, Natividad** [ES/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB). **FORSTER, A.** [GB/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB).

(74) Agents: **MASCHIO, Antonio** et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR GENERATING IMMUNOGLOBULIN GENES

(57) Abstract: The invention provides a method for preparing a double stranded nucleic acid which encodes an immunoglobulin, comprising the steps of: (a) providing a set of three or more overlapping oligonucleotides which anneal to form the + and - strands of a nucleic acid which encodes at least part of an immunoglobulin variable domain; (b) annealing the oligonucleotides; (c) replicating the + and - strands of the nucleic acid formed from the annealed oligonucleotides; and (d) inserting the nucleic acid into an expression vector.

WO 2004/046189 A2

Method for Generating Immunoglobulin Genes

Field of the Invention

The present invention relates to the *de novo* synthesis of immunoglobulin genes and the generation of libraries of diversified immunoglobulin genes. In particular, the invention describes the assembly of antibody genes from oligonucleotides and the footprint mutagenesis of framework and/or CDR regions to generate diversified antibody libraries.

Introduction

Nucleic acids for the production of immunoglobulins by recombinant DNA technology have, in the prior art, been isolated by cloning from natural sources, such as from mouse or human tissues. See, for example, European Patent Application 0 368 684 (MRC/Winter), which describes the cloning of human antibody genes and the construction of libraries of human antibodies. Whilst these methods are suitable for the generation of large libraries of antibody genes based on natural antibodies, such as human antibodies, they do not address the creation of libraries of immunoglobulins not normally seen in nature. Moreover, the techniques for the introduction of diversity into cloned libraries rely on random or semi-random introduction of mutations, which is an inefficient process and results in the production of very large libraries – up to 10^{14} members in size – containing a large number of inoperative mutants which are incapable of binding to any target.

In nature, antibodies operate are secreted by antibody-producing cells and operate extracellularly. However, in recent years the use of antibodies intracellularly, in the cytoplasm and/or the nucleus of cells, has presented a number of advantages in biological ant therapeutic applications. Intracellular antibodies are not produced naturally, and hence there is no natural source of intracellular antibody genes which may be used as a basis for cloning and library preparation in the conventional manner.

Intracellular antibodies or intrabodies are antibody fragments that are used inside cells for interaction with target antigens and for interference with cellular function or in some cases to mediate cell killing following antigen binding. Intrabodies have particular promise in the area of functional genomics where the genome sequence projects are

generating a plethora of open reading frames (Ors) for which no functional data are available. Intrabodies have a role in defining these functions, especially where protein interactions have been defined. In therapeutics, the use of intracellular antibodies for functional ablation has been described and should be an invaluable format for disease-specific reagents.

Intracellular antibodies are typically formulated as single chain Fv (scFv) fragments which comprise immunoglobulin variable (V) regions of heavy (H) and light (L) chains held together by a short linker REF. In the prior art, antigen-specific hybridomas have been used as a source of antibody genes from which scFv have been made for in cell expression as intrabodies, and many successes have been reported in which cellular phenotypes have been obtained due to scFv-antigen binding REF. Conversion of hybridoma antibodies into intracellular antibody fragments is laborious as this strategy requires an antigen-specific hybridoma from which the scFv derivative must be active in the intracellular milieu (which is a reducing environment).

Several different methods have been used to directly develop intrabodies without the need to go through hybridomas. These include genetic screening for intrabody-antigen interaction REF and use of fixed scFv frameworks for intrabody libraries REF. In the former approach, the intracellular antibody capture (IAC) technology REF facilitated the identification of consensus frameworks comprising residues from V_H and V_L which are most commonly found in selected intracellular antibodies. When intracellular antibodies based on these scaffolds are expressed in mammalian cells, they were found to be soluble, well expressed and functionally efficient REF.

Since intracellular antibodies appear to require a very specific scaffold structure, the cloning of antibody genes as a basis for constructing libraries of intrabodies is not feasible. Thus, alternative methods are required for the construction of intracellular antibody libraries.

Summary of the Invention

In our copending UK patent applications entitled "Intracellular Antibodies" and "Anti-activated RAS antibodies", filed on even date herewith, we have demonstrated that the IAC consensus frameworks can be used to convert poor intracellular antibodies into efficient ones by mutating framework residues to the IAC consensus whilst leaving the complementarity determining regions (CDR) intact. In addition, intrabody libraries have been made from consensus frameworks and these can be directly screened for intracellular binders.

We show herein that it is possible to build an intrabody library with only the knowledge of the intracellular antibody consensus sequence, without resorting to any pre-existing antibody gene clones. Two methods are described to achieve this goal. The first is *de novo* antibody gene synthesis in which consensus scFv sequences are used to generate oligonucleotides for gene synthesis and the second is the use of these cloned intracellular antibody genes for CDR or framework diversification using an approach called footprint mutagenesis.

The methods of the invention are widely applicable to the construction of immunoglobulin genes and the diversification of immunoglobulin libraries, for example for the isolation of antibodies having a desired specificity and/or affinity maturation of existing antibody clones. The invention is useful whether the immunoglobulins are artificial and cannot be cloned, such as intrabodies, or are natural and can be cloned. The method of the invention is faster and less laborious than cloning-based methods, and provides leaner, more focussed libraries with fewer inoperative or redundant members.

According to a first aspect, therefore, there is provided a method for preparing a double stranded nucleic acid which encodes an immunoglobulin, comprising the steps of:

(a) providing a set of three or more overlapping oligonucleotides which anneal to form the + and – strands of a nucleic acid which encodes at least part of an immunoglobulin variable domain;

(b) annealing the oligonucleotides;

(c) replicating the + and – strands of the nucleic acid formed from the annealed oligonucleotides; and

(d) inserting the nucleic acid into an expression vector.

Preferably, the oligonucleotide set encodes an entire V_H or V_L domain, or both. In an advantageous embodiment, the set encodes a scFv molecule comprising both V_H and V_L domains, and a linker coupling said domains. In a further embodiment, the oligonucleotides encode a single domain antibody or DAb.

Oligonucleotides may also be provided which encode one or more constant region domains, or other effector groups or labels, which are attached to the immunoglobulin molecule. For example, the oligonucleotides may encode all or part of an antibody Fc region, a label such as GFP, or an effector group such as a toxin. The method may moreover be used to construct multiple domain immunoglobulins, such as bispecific antibodies.

Thus, where the nucleic acid molecule encodes "at least part" of an immunoglobulin variable domain, it will be understood that the nucleic acid molecule preferably encodes an entire variable domain, and may encode additional domains or further polypeptides. Advantageously, the nucleic acid molecule encodes the entirety of the gene product which it is desired to produce; thus, it may encode an entire scFv, and entire DAb, an entire V_H or V_L domain, or any of these which has been modified by addition of one or more constant region domains, effector groups or labels.

In a preferred aspect, the invention further comprises the steps of:

(e) amplifying the nucleic acid molecule encoding the immunoglobulin variable domain using a first set of at least four primers, of which two primers are overlapping, and wherein one of the overlapping primers consists of a plurality of different primers, each of which comprises a mutagenic region which generates a mutation in a part of the nucleic acid sequence;

(f) purifying the amplification products thus obtained;

(g) assembling the amplification products in a further amplification reaction using a second set of primers which encompass the entire nucleic acid molecule which encodes the immunoglobulin variable domain; and

(h) inserting the further amplification product into an expression vector.

Advantageously, the nucleic acid molecule and further amplification product which is finally obtained after the mutagenesis procedure encode the gene product which it is desired to obtain in its entirety, as specified above.

Preferably, steps (e) to (g) are repeated, using different primers, to mutate different regions of the nucleic acid molecule and thus generate a library of diversified nucleic acid molecules which encompass mutations in a plurality of defined areas.

In an advantageous embodiment, primers are used which comprise a multiple codon equivalence of randomised or semi-randomised sequence. For instance, these can be designed to cover a CDR, thus providing the ability to create a diversified CDR and thus a library of antibody molecules having a diversified CDR. Moreover, multiple primer sets may be used simultaneously to diversify two or more separate regions, such as for example two or more CDRs, in order to provide a further diversified library.

Where two or more primers are used, for example to diversify two or more regions in the immunoglobulin chain, the PCR products generated advantageously overlap. The overlap is preferably in a region complementary to both primers. Advantageously the overlap does not include a diversified region.

Multiple codon equivalence may extend over two to twelve or more codons; preferably, it extends to between three and ten codons.

Where separate PCR products are generated they are assembled using flanking primers which amplify the whole region of interest; the flanking primers used may be the same and/or different to the diversification primers used, but preferably comprise no diversified sequence.

In a further aspect, the invention provides a method for preparing a library of nucleic acids encoding a diversified immunoglobulin, comprising the steps of:

(a) amplifying a nucleic acid molecule encoding the immunoglobulin using a first set of at least four primers, of which two primers are overlapping, and wherein one of the overlapping primers consists of a plurality of different primers, each of which comprises a mutagenic region which generates a mutation in a part of the nucleic acid sequence;

(b) purifying the amplification products thus obtained;

(c) assembling the amplification products in a further amplification reaction using a second set of primers which encompass the entire nucleic acid molecule which encodes the immunoglobulin; and

(d) inserting the assembled amplification product into an expression vector.

The method of this aspect of the invention provides a means to generate diversified immunoglobulin libraries without recourse to phage display or other expression display methodologies.

The invention is particularly useful for the production of antibodies and antibody libraries, in particular libraries of antibody variable domains. These may be configured as scFv, Fabs or other fragments. Particularly preferred are single domain antibodies or DAbs.

Advantageously, the invention is useful for the generation of intracellular antibodies, such as intracellular scFv or intracellular DAbs.

The invention accordingly provides a library of nucleic acids prepared according to the method of the invention. Advantageously, it is a library of intracellular antibodies.

Libraries of intracellular antibodies are advantageously generated by diversifying one, two or three CDRs on the intracellular antibody consensus framework described in our copending international patent application PCT/GB02/003512. This framework provides a basis on which intracellular antibodies may be based and reliably function in an intracellular environment; the use of the methods of the present invention allows diversification of the CDRs on the basis of this framework and thus the generation

libraries of intracellular immunoglobulins having low redundancy and reduced incidence of intracellular incompatibility.

Brief Description of the Figures

Figure 1. *De novo* antibody gene synthesis

A. Flow diagram of *de novo* antibody gene synthesis. Step 1. Oligonucleotides corresponding to both strands of the desired antibody fragment (in this case an scFv but could be VH or VL alone) are mixed, annealed and ligated. Step 2. PCR amplification of whole scFv is achieved using flanking primers (conseSFI + conseNOT) carrying Sfi1 or Not1 sites. Step 3. The PCR product is cleaved with Sfi1 and Not1 and cloned into a compatible vector, in this case pEF-VP16. This vector was constructed from pEF/myc/cyto (Invitrogen) by addition of the VP16 transcriptional activation domain and mutation of the Sfi1 site for compatibility with most scFv cloned sequences (see methods).

B. Sequence of hybrid scFv and location of oligonucleotides used for gene synthesis. The design of scFvconR4 was using the intracellular antibody capture consensus framework sequence (Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, T. Y. G. and Rabbitts, T. H. (2002) *J. Mol. Biol.*, **317,** 85-94) with the VH and VLλ CDR sequences from anti-βgal scFv R4 (Martineau, P., Jones, P. and Winter, G. (1998) *J Mol Biol*, **280,** 117-127). CDR regions are in yellow.

C. Mammalian cell reporter assay for scFv intrabody activity. CHO-CD4 cells (Fearon, E. R., Finkel, T., Gillison, M. L., Kennedy, S. P., Casella, J. F., Tomaselli, G. F., Morrow, J. S. and Dang, C. V. (1992) *Proc. Natl. Acad. Sci. USA*, **89,** 7958-7962) were transfected as indicated and stimulation of CD4 surface expression was measured using a FACS with anti-human CD4 antibody and FITC-anti-mouse antibody. Cells were transfected with combinations of plasmids encoding DBD-βgal (lacZ), scFvR4-VP16 (scFvR4), DBD-RASG12V (RASG12V) and/or scFvconR4-VP16 (scFvconR4) as indicated.

Figure 2. Footprint mutagenesis of scFv frameworks

Specific framework residues of the intrabody scFv33 (see our copending UK patent application "Anti-activated RAS antibodies" filed on even date herewith) was mutated to those of scFvI21 to yield scFvI21R33 (16) by stepwise footprint mutagenesis.

A. At each step of mutagenesis, the scFv template, cloned in pEF-VP16, was PCR amplified using a fixed primer (EFFP2 or VP162R) together with a mutant primer (indicated as --^--) at a specific position; this yields two PCR fragments which are assembled with EFFP + VP162R primers and cloned into pEF-VP16 for the next round of mutagenesis.

B. Nucleotide and derived protein sequences of scFv33, indicating the amino acid residues mutated in the stepwise footprint mutagenesis to scFvI21R33. The PCR primers are shown above (red; forward primers) and below (blue; reverse primers) the template sequence. The CDR regions are highlighted in yellow and linker between VH and VL in grey. The substituted amino acid residues are shown in italics.

Figure 3. Preparation of scFv intrabodies with randomised CDR3 using footprint mutagenesis

The intracellular antibody scFvA25, recognising the BCR-ABL oncogenic protein (Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, T. Y. G. and Rabbitts, T. H. (2002) *J. Mol. Biol.*, **317,** 85-94), was used as a template for a diversified library with randomised mutations of the VH CDR3 region.

A. The scFvA25 was cloned into the pEF-VP16 vector and two footprint mutagenesis PCR reactions carried out with primers EFFP2 + A25C3B (in which the central region of the primer has n=3, 6 or 10 to randomise amino acid residues in CDR3 of the VH domain). The two PCR products were mixed, assembled and cloned into the yeast vector pVP16*. CDRs are highlighted in yellow.

B. The sequence of A25 VH CDR3 region (highlighted in yellow) and PCR primers A25CD3F + $A25C3B_N$.

C. The DNA sequences of randomly selected clones from each 1st PCR were obtained and the derived VH CDR3 protein sequences of these clones are shown (highlighted in yellow).

Figure 4. Diversification of VH CDR1, 2 and 3 using footprint mutagenesis

A. Footprint mutagenesis to generate an V-region segment with randomised CDR regions. The template illustrated is a VH segment and the CDR2 and CDR3 regions are mutated as shown. Step 1. Two PCR reactions were carried out with EFFP2 + CDR2R (which randomises CDR2 as shown in B)) and CDR2F + CDR3R (which randomises

CDR3 as shown in B)). Step 2. The two reaction products were assembled into a complete VH sequence using EFFP2 + JH5R flanking primers and this in turn amplified with partially nested primers EFFP + NOTVHJR1 to incorporate Sfi1 and Not1 restriction sites. Step 3. Cloning PCR products into yeast pVP16* (Visintin, M., Tse, E., Axelson, H., Rabbitts, T. H. and Cattaneo, A. (1999) *Proc. Natl. Acad. Sci. USA*, **96,** 11723-11728). CDRs are highlighted in yellow. Forward primers are shown in red and reverse primers in blue.

B. This illustrates a VH domain depicting framework regions (FR) and complementarity determining regions (CDR; highlighted in yellow) with the PCR oligonucleotide sequences used mutagenesis. For first round mutagenesis, CDR2 and CDR3 were simultaneously changed as indicated. For second round mutagenesis, CDR1 was changed as indicated.

C. Derived protein sequences of the CDR regions of five selected clones made by mutation of CDR2 + CDR3 compared with the CDR1/2/3 sequences (top line) of the canonical consensus VH sequence (CDR2/3) or five selected clones made by randomisation of CDR1 from a library of 3.4 X 10^6 clones with mutated CDR2 + CDR3 (CDR1/2/3); CDR residues are highlighted in grey. The regions randomised in the PCR oligonucleotides is shown in the second line, with mutant residues highlighted in grey.

Figure 5 shows the Alignment of derived protein sequences of intracellular scFv.
The nucleotide sequences of the scFv were obtained and the derived protein translations (shown in the single letter code) were aligned. The complementarity determining regions (CDR) are shaded. Framework residues for SEQ no 1 to 40 are those which are underlined. The consensus sequence at a specific position was calculated for the most frequently occurring residue but only conferred if a residue occurred greater than 5 times at that position.

A. Sequences of VH and VL from anti-BCR (designated as B3-B89) and anti-ABL (designated as A5-A32). The combined consensus (Con) of the anti-BCR and ABL ICAbs is indicated compared with the subgroup consensuses forVH3 and V_KI from the Kabat database .

- Represents sequence identity with the intracellular antibody binding V_H or V_L consensus (SEQ. ID. No. 3 and SEQ. ID. No. 4 respectively)

. represents gaps introduced to optimise alignment

B. A sequence comparison of randomly obtained scFv obtained from the unselected phage display library. The consensuses obtained from the randomly isolated scFv (rcH and rcL) are indicated.

- represents gaps introduced to optimise alignment

X represents positions at which no consensus could be assigned.

Detailed Description of the Invention

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art (e.g., in cell culture, molecular genetics, nucleic acid chemistry, hybridisation techniques and biochemistry). Standard techniques are used for molecular, genetic and biochemical methods. See, generally, Sambrook et al., Molecular Cloning: A Laboratory Manual, 2d ed. (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. and Ausubel et al., Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc.; as well as Guthrie et al., Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology, Vol. 194, Academic Press, Inc., (1991), PCR Protocols: A Guide to Methods and Applications (Innis, et al. 1990. Academic Press, San Diego, Calif.), McPherson et al., PCR Volume 1, Oxford University Press, (1991), Culture of Animal Cells: A Manual of Basic Technique, 2nd Ed. (R. I. Freshney. 1987. Liss, Inc. New York, N.Y.), and Gene Transfer and Expression Protocols, pp. 109-128, ed. E. J. Murray, The Humana Press Inc., Clifton, N.J.). These documents are incorporated herein by reference.

Definitions

A **nucleic acid**, as referred to herein, is a sequence of nucleotides which is advantageously a DNA sequence. The nucleotides may be natural or synthetic, or a mixture of the two; the nucleic acids may be linear or circular in form as required.

Immunoglobulins, according to the present invention include members of the immunoglobulin superfamily, a family of polypeptides which comprise the immunoglobulin fold characteristic of antibody molecules, which contains two β sheets

and, usually, a conserved disulphide bond. Members of the immunoglobulin superfamily are involved in many aspects of cellular and non-cellular interactions *in vivo*, including widespread roles in the immune system (for example, antibodies, T-cell receptor molecules and the like), involvement in cell adhesion (for example the ICAM molecules) and intracellular signalling (for example, receptor molecules, such as the PDGF receptor). The present invention is preferably relates to single domain immunoglobulins derived from all immunoglobulin superfamily molecules which are capable of binding to target molecules. Preferably, the present invention relates to antibody single domains, in particular heavy chain variable (V_H) domains. Single domain immunoglobulins are free of complementary domains, that is are not associated with other binding domains which, in nature or otherwise, may associate with the single domain to form a single composite binding site for a target. Specifically, V_H domains are not in the presence of complementary V_L domains in the single domains of the invention. However, further domains, such as antibody constant region domains, may be but need not be present.

Antibodies, as used herein, refers to complete antibodies or antibody fragments capable of binding to a selected target, and including Fv, scFv, Fab' and $F(ab')_2$, monoclonal and polyclonal antibodies, engineered antibodies including chimeric, CDR-grafted and humanised antibodies, and artificially selected antibodies produced using phage display or alternative techniques. Antibodies may be or be based on of any naturally-occurring antibody type, including IgG, IgE, IgA, IgD and IgM. Single domains, such as V_H domains, may be derived from any such antibody.

The **framework region** of an immunoglobulin heavy and/or light chain variable domain has a particular 3 dimensional conformation characterised by the presence of an immunoglobulin fold. Certain amino acid residues present in the variable domain are responsible for maintaining this characteristic immunoglobulin domain core structure. These residues are known as framework residues and tend to be highly conserved. The framework supports the CDRs of an antibody.

CDR (complementarity determining region) of an immunoglobulin molecule heavy and/or light chain variable domain describes those amino acid residues which are not framework region residues and which are contained within the hypervariable loops of the

variable regions. These hypervariable loops are directly involved with the interaction of the immunoglobulin with the ligand. Residues within these loops tend to show less degree of conservation than those in the framework region.

Intracellular means inside a cell, and the present invention is directed to those immunoglobulins which will bind to ligands/targets selectively within a cell. The cell may be any cell, prokaryotic or eukaryotic, and is preferably selected from the group consisting of a bacterial cell, a yeast cell and a higher eukaryote cell. Most preferred are yeast cells and mammalian cells. As used herein, therefore, "intracellular" immunoglobulins and targets or ligands are immunoglobulins and targets/ligands which are present within a cell. In addition the term 'Intracellular' refers to environments which resemble or mimic an intracellular environment. Thus, "intracellular" may refer to an environment which is not within the cell, but is *in vitro*. For example, the method of the invention may be performed in an *in vitro* transcription and/or translation system, which may be obtained commercially, or derived from natural systems.

Consensus frameworks in the context of the present invention refers to the consensus sequences of those V_H and V_L chains from immunoglobulin molecules which can bind selectively to a ligand in an intracellular environment. The residue which is most common in any one given position, when the sequences of those immunoglobulins which can bind intracellularly are compared is chosen as the consensus residue for that position. The consensus sequence is generated by comparing the residues for all the intracellularly binding immunoglobulins, at each position in turn, and then collating the data.

Oligonucleotides are nucleic acids composed of a plurality of nucleotides. Advantageously, they are useful in assembly of larger nucleic acids as described herein. No limit in length is intended to be implied, and oligonucleotides may be short – 3 to 10 nucleotides long – to long, 1000 nucleotides long and more. Preferably, oligonucleotides are synthesised.

Double stranded nucleic acids possess + and – strands, which equates to a coding strand and a non-coding strand in coding sequences. The + and – strands are complementary, and anneal to form a nucleic acid duplex.

Annealing is the process by which complementary nucleic acids hybridise to each other to form a double stranded nucleic acid. Nucleic acids do not need to be 100% complementary to anneal to each other.

Effector domains are polypeptides which are capable of exerting a chemical or biological function. For example, antibody Fc regions comprise effector domains, which can be attached to scFv or Dab antibodies. Labels are advantageously protein labels, such as luminescent labels (for example GFP) or antigenic epitopes.

Nucleic acids according to the invention are **amplified** using any available amplification technique. "Amplification" refers to the increase in the number of copies of a particular nucleic acid fragment (or a portion of this) resulting from an enzymatic chain reaction (such as a polymerase chain reaction, a ligase chain reaction, or a self-sustained sequence replication). Preferably, the amplification according to our invention is an exponential amplification, as exhibited by for example the polymerase chain reaction (PCR).

Primers are nucleic acid molecules which are used to prime amplification reactions. They are complementary to a part of the nucleic acids which it is desired to amplify, and allow the amplification process to begin at the point at which they anneal. Primers are **mutagenic** if they insert, into a nucleic acid to be amplified, a mutation due to a base pair mismatch which is tolerated in the primer annealing to the nucleic acid but carried over to the amplified product.

Amplification products are **assembled**, in the context of the present invention, by ligation or otherwise to produce a composite nucleic acid comprising more than one amplification product. In a preferred embodiment, the amplification products are assembled by PCR or another reaction involving nucleic acid replication which employs primers located at either end of the desired assembled molecule, thus directing replication and/or amplification of the full length assembled nucleic acid.

Mutagenic primers may comprise **randomised** and/or **semi-randomised** sequence. Randomised sequence occurs at a position in the primer where any nucleotide, A, C, G or

T, may appear. This is usually denoted by the letter N in the sequence. Thus, a codon represented as 'NNN' may encode any amino acid, or a be a nonsense or stop codon. Semi-randomised sequences are restricted in the degree of randomisation, such that not any nucleotide may appear at any position. For instance, the nature of the third base in any codon may be restricted, to avoid the incidence of stop codons. The randomised sequence extends over a number of codons, which is the **codon equivalence** of the randomisation. A multiple codon equivalence indicates that two or more codons are (semi-) randomised. Preferably, between 3 and 12 codons are (semi-) randomised; advantageously, 6 to 10 codons are (semi-) randomised.

Mutagenesis using randomised sequences as described leads to **diversification** of the sequence of the nucleic acids and/or primers according to the invention, and thus the generation of libraries of diversified polypeptides by expression of the nucleic acids. **Expression** is the transcription and/or translation of nucleic acid into the gene product it encodes; in the context of the present invention, that gene product is a polypeptide. Thus, expression is the transcription and/or translation of nucleic acids to form polypeptides. The polypeptides are advantageously immunoglobulins.

An **expression vector** is a nucleic acid which comprises a coding sequence and the sequences necessary for that coding sequence to be expressed, as defined herein. Typically, an expression vector will be a plasmid, and will comprise one or more origins of replication, a promoter and optionally enhancer sequences to direct transcription of the coding sequences, and optionally one or more markers which allow the vector or cells containing the vector to be identified and/or selected for.

Isolation, as referred to herein, is the purification of the desired substance from one or more undesired components with which it is associated. Thus, isolation of nucleic acids according to the invention indicates that the nucleic acids are purified from unreacted nucleotides, primers, enzymes and other reaction components with which they are associated with after a given reaction. The degree of purification need not be complete purification; it suffices to isolate the desired nucleic acids sufficiently to allow them to be used in further processes and/or reactions.

AMPLIFICATION

The methods of the invention involve the templated replication and/or amplification of desired nucleic acids. "Amplification" refers to the increase in the number of copies of a particular nucleic acid fragment (or a portion of this) resulting from an enzymatic chain reaction (such as a polymerase chain reaction, a ligase chain reaction, or a self-sustained sequence replication. Preferably, the amplification is an exponential amplification, as exhibited by for example the polymerase chain reaction.

Many target and signal amplification methods have been described in the literature. See, for example, general reviews of these methods in Landegren, U., et al., Science 242:229-237 (1988) and Lewis, R., Genetic Engineering News 10:1, 54-55 (1990). These amplification methods may be used in the methods of the invention, and include polymerase chain reaction (PCR), PCR in situ, ligase amplification reaction (LAR), ligase hybridisation, Qbeta bacteriophage replicase, transcription-based amplification system (TAS), genomic amplification with transcript sequencing (GAWTS), nucleic acid sequence-based amplification (NASBA) and *in situ* hybridisation. The use of PCR is preferred.

Polymerase Chain Reaction (PCR)

PCR is a nucleic acid amplification method described *inter alia* in U.S. Pat. Nos. 4,683,195 and 4,683,202. PCR consists of repeated cycles of DNA polymerase generated primer extension reactions. The target DNA is heat denatured and two oligonucleotides, which bracket the target sequence on opposite strands of the DNA to be amplified, are hybridised. These oligonucleotides become primers for use with DNA polymerase. The DNA is copied by primer extension to make a second copy of both strands. By repeating the cycle of heat denaturation, primer hybridisation and extension, the target DNA can be amplified a million fold or more in about two to four hours. An advantage of PCR is that it increases sensitivity by amplifying the amount of target DNA by 1 million to 1 billion fold in approximately 4 hours. In the context of the present invention, PCR is used to amplify desired gene products, to assemble amplification products of other PCR reactions into full-length nucleic acids, and to introduce mutations into nucleic acids using primers which comprises randomised sequences.

Reverse transcriptase-PCR

RT-PCR is used to amplify RNA targets. In this process, the reverse transcriptase enzyme is used to convert RNA to complementary DNA (cDNA), which can then be amplified using PCR. This method has proven useful for the detection of RNA viruses.

The methods of the invention may employ RT-PCR. Thus, the nucleic acid encoding the immunoglobulin may be provided in the form of RNA. This RNA could be generated *in vivo* in bacteria, mammalian cells, yeast etc, and may for example be the transcription product of endogenous immunoglobulin genes.

Ligation Amplification (LAR/LAS)

Ligation amplification reaction or ligation amplification system uses DNA ligase and four oligonucleotides, two per target strand. This technique is described by Wu, D. Y. and Wallace, R. B. (1989) Genomics 4:560. The oligonucleotides hybridise to adjacent sequences on the target DNA and are joined by the ligase. The reaction is heat denatured and the cycle repeated. The opposite strand may be copied by any replicase enzyme.

Qβ Replicase

In this technique, RNA replicase for the bacteriophage Qβ, which replicates single-stranded RNA, is used to amplify the target DNA, as described by Lizardi et al. (1988) Bio/Technology 6:1197. First, the target DNA is hybridised to a primer including a T7 promoter and a Qβ 5' sequence region. Using this primer, reverse transcriptase generates a cDNA connecting the primer to its 5' end in the process. These two steps are similar to the TAS protocol. The resulting heteroduplex is heat denatured. Next, a second primer containing a Qβ 3' sequence region is used to initiate a second round of cDNA synthesis. This results in a double stranded DNA containing both 5' and 3' ends of the Qβ bacteriophage as well as an active T7 RNA polymerase binding site. T7 RNA polymerase then transcribes the double-stranded DNA into new RNA, which mimics the Qβ. After extensive washing to remove any unhybridized probe, the new RNA is eluted from the target and replicated by Qβ replicase. The latter reaction creates 10^7 fold amplification in approximately 20 minutes. Significant background may be formed due to minute amounts of probe RNA that is non-specifically retained during the reaction.

Other Amplification Techniques

Alternative amplification technology may be exploited in the present invention. For example, rolling circle amplification (Lizardi *et al.*, (1998) Nat Genet 19:225) is an amplification technology available commercially (RCAT™) which is driven by DNA polymerase and can replicate circular oligonucleotide probes with either linear or geometric kinetics under isothermal conditions.

In the presence of two suitably designed primers, a geometric amplification occurs via DNA strand displacement and hyperbranching to generate 10^{12} or more copies of each circle in 1 hour.

A further technique, strand displacement amplification (SDA; Walker *et al.*, (1992) PNAS (USA) 80:392) begins with a specifically defined sequence unique to a specific target. But unlike other techniques which rely on thermal cycling, SDA is an isothermal process that utilises a series of primers, DNA polymerase and a restriction enzyme to exponentially amplify the unique nucleic acid sequence.

SDA comprises both a target generation phase and an exponential amplification phase. In target generation, double-stranded DNA is heat denatured creating two single-stranded copies. A series of specially manufactured primers combine with DNA polymerase (amplification primers for copying the base sequence and bumper primers for displacing the newly created strands) to form altered targets capable of exponential amplification. The exponential amplification process begins with altered targets (single-stranded partial DNA strands with restricted enzyme recognition sites) from the target generation phase. An amplification primer is bound to each strand at its complimentary DNA sequence. DNA polymerase then uses the primer to identify a location to extend the primer from its 3' end, using the altered target as a template for adding individual nucleotides. The extended primer thus forms a double-stranded DNA segment containing a complete restriction enzyme recognition site at each end.

A restriction enzyme is then bound to the double stranded DNA segment at its recognition site. The restriction enzyme dissociates from the recognition site after having cleaved

only one strand of the double-sided segment, forming a nick. DNA polymerase recognises the nick and extends the strand from the site, displacing the previously created strand. The recognition site is thus repeatedly nicked and restored by the restriction enzyme and DNA polymerase with continuous displacement of DNA strands containing the target segment.

Each displaced strand is then available to anneal with amplification primers as above. The process continues with repeated nicking, extension and displacement of new DNA strands, resulting in exponential amplification of the original DNA target.

EXPRESSION SYSTEMS

Methods which are well known to those skilled in the art are used to construct expression vectors containing sequences encoding immunoglobulins according to the invention and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989; Molecular Cloning, A Laboratory Manual, ch. 4, 8, and 16-17, Cold Spring Harbor Press, Plainview, N.Y.) and Ausubel, F. M. et al. (1995 and periodic supplements; Current Protocols in Molecular Biology, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.).

A variety of expression vector/host systems may be utilised to contain and express sequences encoding immunoglobulins. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector (i.e., enhancers, promoters, and 5' and 3' untranslated regions) which interact with

host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilised, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPorT1 plasmid (GIBCO/BRL), and the like, may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding an immunoglobulin, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected. For example, vectors which direct high level expression of fusion proteins may be used. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding the immunoglobulin may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β-galactosidase so that a hybrid protein is produced, pIN vectors (Van Heeke, G. and S. M. Schuster (1989) J. Biol. Chem. 264:5503-5509), and the like. pGEX vectors (Promega, Madison, Wis.) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. For reviews, see Ausubel (supra) and Grant et al. (1987; Methods Enzymol. 153:516-544).

In cases where plant expression vectors are used, the expression of sequences encoding the immunoglobulin may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) EMBO J. 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews. (See, for example, Hobbs, S. or Murry, L. E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, N.Y.; pp. 191-196.).

An insect system may also be used to express the immunoglobulin. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding the immunoglobulin may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the immunoglobulin gene will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which GPR54 polypeptide may be expressed. (Engelhard, E. K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227.)

In mammalian host cells, a number of viral-based expression systems may be utilised. In cases where an adenovirus is used as an expression vector, sequences encoding the immunoglobulin may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the immunoglobulin in infected host cells. (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding the immunoglobulin. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the immunoglobulin and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular cell system used, such as those described in the literature. (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC, Bethesda, Md.) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is preferred. For example, cell lines capable of stably expressing the immunoglobulin can be

transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase genes (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase genes (Lowy, I. et al. (1980) Cell 22:817-23), which can be employed in tk^- or apr^- cells, respectively. Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, dhfr confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14); and als or pat confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilise indole in place of tryptophan, or hisD, which allows cells to utilise histinol in place of histidine. (Hartman, S. C. and R. C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51.) Recently, the use of visible markers has gained popularity with such markers as anthocyanins, β-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (Rhodes, C. A. et al. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding the immunoglobulin is inserted within a marker gene sequence, transformed cells containing sequences encoding the immunoglobulin can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding the immunoglobulin

under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which express the immunoglobulin may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA--DNA or DNA-RNA hybridizations and immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

The presence of polynucleotide sequences encoding the immunoglobulin can be detected by DNA-DNA or DNA-RNA hybridisation or amplification using probes or fragments or fragments of polynucleotides encoding the immunoglobulin. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding the immunoglobulin, as described above, to detect transformants containing DNA or RNA encoding the immunoglobulin.

A variety of protocols for detecting and measuring the expression of immunoglobulins are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). These and other assays are well described in the art, for example, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, Section IV, APS Press, St Paul, Minn.) and in Maddox, D. E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labelled hybridisation or PCR probes for detecting sequences related to polynucleotides encoding immunoglobulins include oligolabeling, nick translation, end-labelling, or PCR amplification using a labelled nucleotide. Alternatively, the sequences encoding the immunoglobulin, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesise RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labelled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by

Pharmacia & Upjohn (Kalamazoo, Mich.), Promega (Madison, Wis.), and U.S. Biochemical Corp. (Cleveland, Ohio). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding the immunoglobulin may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be located in the cell membrane, secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode the immunoglobulin may be designed to contain signal sequences which direct secretion of the immunoglobulin through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding the immunoglobulin to nucleotide sequences encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilised metals, protein A domains that allow purification on immobilised immunoglobulin, and the domain utilised in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.), between the purification domain and the immunoglobulin-encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing the immunoglobulin and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on immobilised metal ion affinity chromatography (IMIAC; described in Porath, J. et al. (1992) Prot. Exp. Purif. 3: 263-281), while the enterokinase cleavage site provides a means for purifying GPR54 from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; DNA Cell Biol. 12:441-453).

IMMUNOGLOBULINS

Immunoglobulin molecules, according to the present invention, refer to members of the immunoglobulin superfamily, a family of polypeptides which comprise the immunoglobulin fold characteristic of antibody molecules, which contains two β sheets and, usually, a conserved disulphide bond. Members of the immunoglobulin superfamily are involved in many aspects of cellular and non-cellular interactions *in vivo*, including widespread roles in the immune system (for example, antibodies, T-cell receptor molecules and the like), involvement in cell adhesion (for example the ICAM molecules) and intracellular signalling (for example, receptor molecules, such as the PDGF receptor). The present invention is applicable to all immunoglobulin superfamily molecules which are capable of binding to target molecules. Preferably, the present invention relates to antibodies.

Antibodies, as used herein, refers to complete antibodies or antibody fragments capable of binding to a selected target, and including Fv, scFv, Fab' and $F(ab')_2$, monoclonal and polyclonal antibodies, engineered antibodies including chimeric, CDR-grafted and humanised antibodies, and artificially selected antibodies produced using phage display or alternative techniques. Small fragments, such as Fv and scFv, possess advantageous properties for diagnostic and therapeutic applications on account of their small size and consequent superior tissue distribution. Preferably, the antibody is a single chain antibody or scFv.

The antibodies according to the invention are especially indicated for diagnostic and therapeutic applications. Accordingly, they may be altered antibodies comprising an effector protein such as a toxin or a label. Especially preferred are labels which allow the imaging of the distribution of the antibody in vivo. Such labels may be radioactive labels or radioopaque labels, such as metal particles, which are readily visualisable within the body of a patient. Moreover, they may be fluorescent labels or other labels which are visualisable on tissue samples removed from patients. Effector groups may be added during the synthesis of the antibodies by the method of the present invention, or afterwards.

Recombinant DNA technology may be used to produce the antibodies of then invention according to established procedure, in bacterial or preferably mammalian cell culture. The selected cell culture system preferably secretes the antibody product.

Multiplication of mammalian host cells *in vitro* is carried out in suitable culture media, which are the customary standard culture media, for example Dulbecco's Modified Eagle Medium (DMEM) or RPMI 1640 medium, optionally replenished by a mammalian serum, e.g. foetal calf serum, or trace elements and growth sustaining supplements, e.g. feeder cells such as normal mouse peritoneal exudate cells, spleen cells, bone marrow macrophages, 2-aminoethanol, insulin, transferrin, low density lipoprotein, oleic acid, or the like. Multiplication of host cells which are bacterial cells or yeast cells is likewise carried out in suitable culture media known in the art, for example for bacteria in medium LB, NZCYM, NZYM, NZM, Terrific Broth, SOB, SOC, 2 x YT, or M9 Minimal Medium, and for yeast in medium YPD, YEPD, Minimal Medium, or Complete Minimal Dropout Medium.

In vitro production provides relatively pure antibody preparations and allows scale-up to give large amounts of the desired antibodies. Techniques for bacterial cell, yeast or mammalian cell cultivation are known in the art and include homogeneous suspension culture, e.g. in an airlift reactor or in a continuous stirrer reactor, or immobilised or entrapped cell culture, e.g. in hollow fibres, microcapsules, on agarose microbeads or ceramic cartridges.

The foregoing, and other, techniques are discussed in, for example, Harlow and Lane, Antibodies: a Laboratory Manual, (1988) Cold Spring Harbor, incorporated herein by reference. Techniques for the preparation of recombinant antibody molecules is described in the above reference and also in, for example, EP 0623679, EP 0368684 and EP 0436597, which are incorporated herein by reference.

The cell culture supernatants are screened for the desired antibodies, preferentially by immunofluorescent staining of cells expressing the desired target by immunoblotting, by an enzyme immunoassay, e.g. a sandwich assay or a dot-assay, or a radioimmunoassay.

For isolation of the antibodies, the immunoglobulins in the culture supernatants or in the ascitic fluid may be concentrated, e.g. by precipitation with ammonium sulphate, dialysis against hygroscopic material such as polyethylene glycol, filtration through selective membranes, or the like. If necessary and/or desired, the antibodies are purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose and/or (immuno-)affinity chromatography, e.g. affinity chromatography with the target molecule or with Protein-A.

Immunoglobulin libraries constructed according to the invention may used in any library selection procedure. Selection protocols for isolating desired members of libraries are known in the art, as typified by phage display techniques. Such systems, in which diverse peptide sequences are displayed on the surface of filamentous bacteriophage have proven useful for creating libraries of antibody fragments (and the nucleotide sequences that encoding them) for the *in vitro* selection and amplification of specific antibody fragments that bind a target antigen. The nucleotide sequences encoding the V_H and V_L regions are linked to gene fragments which encode leader signals that direct them to the periplasmic space of *E. coli* and as a result the resultant antibody fragments are displayed on the surface of the bacteriophage, typically as fusions to bacteriophage coat proteins (e.g., pIII or pVIII). Alternatively, antibody fragments are displayed externally on lambda phage capsids (phagebodies). An advantage of phage-based display systems is that, because they are biological systems, selected library members can be amplified simply by growing the phage containing the selected library member in bacterial cells. Furthermore, since the nucleotide sequence that encode the polypeptide library member is contained on a phage or phagemid vector, sequencing, expression and subsequent genetic manipulation is relatively straightforward.

Methods for the construction of bacteriophage antibody display libraries and lambda phage expression libraries are well known in the art (McCafferty *et al.* (1990) **supra**; Kang *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.*, **88**: 4363; Clackson *et al.* (1991) *Nature*, **352**: 624; Lowman *et al.* (1991) *Biochemistry*, **30**: 10832; Burton *et al.* (1991) *Proc. Natl. Acad. Sci U.S.A.*, **88**: 10134; Hoogenboom *et al.* (1991) Nucleic Acids Res., **19**: 4133; Chang *et al.* (1991) *J. Immunol.*, **147**: 3610; Breitling *et al.* (1991) *Gene*, **104**: 147; Marks *et al.* (1991) supra; Barbas *et al.* (1992) supra; Hawkins and Winter (1992) *J. Immunol.*,

22: 867; Marks *et al.*, 1992, *J. Biol. Chem.*, **267**: 16007; Lerner *et al.* (1992) *Science*, **258**: 1313, incorporated herein by reference).

One particularly advantageous approach has been the use of scFv phage-libraries (Huston *et al.*, 1988, Proc. Natl. Acad. Sci U.S.A., 85: 5879-5883; Chaudhary *et al.* (1990) Proc. Natl. Acad. Sci U.S.A., 87: 1066-1070; McCafferty *et al.* (1990) supra; Clackson *et al.* (1991) supra; Marks *et al.* (1991) supra; Chiswell *et al.* (1992) Trends Biotech., **10**: 80; Marks *et al.* (1992) supra). Various embodiments of scFv libraries displayed on bacteriophage coat proteins have been described. Refinements of phage display approaches are also known, for example as described in WO96/06213 and WO92/01047 (Medical Research Council *et al.*) and WO97/08320 (Morphosys), which are incorporated herein by reference.

Alternative library selection technologies include bacteriophage lambda expression systems, which may be screened directly as bacteriophage plaques or as colonies of lysogens, both as previously described (Huse *et al.* (1989) *Science*, **246**: 1275; Caton and Koprowski (1990) *Proc. Natl. Acad. Sci. U.S.A.*, **87**; Mullinax *et al.* (1990) *Proc. Natl. Acad. Sci. U.S.A.*, **87**: 8095; Persson *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.*, **88**: 2432) and are of use in the invention. Whilst such expression systems can be used to screening up to 10^6 different members of a library, they are not really suited to screening of larger numbers (greater than 10^6 members). Other screening systems rely, for example, on direct chemical synthesis of library members. One early method involves the synthesis of peptides on a set of pins or rods, such as described in WO84/03564. A similar method involving peptide synthesis on beads, which forms a peptide library in which each bead is an individual library member, is described in U.S. Patent No. 4,631,211 and a related method is described in WO92/00091. A significant improvement of the bead-based methods involves tagging each bead with a unique identifier tag, such as an oligonucleotide, so as to facilitate identification of the amino acid sequence of each library member. These improved bead-based methods are described in WO93/06121.

Another chemical synthesis method involves the synthesis of arrays of peptides (or peptidomimetics) on a surface in a manner that places each distinct library member (e.g., unique peptide sequence) at a discrete, predefined location in the array. The identity of

each library member is determined by its spatial location in the array. The locations in the array where binding interactions between a predetermined molecule (e.g., a receptor) and reactive library members occur is determined, thereby identifying the sequences of the reactive library members on the basis of spatial location. These methods are described in U.S. Patent No. 5,143,854; WO90/15070 and WO92/10092; Fodor *et al.* (1991) *Science*, **251**: 767; Dower and Fodor (1991) *Ann. Rep. Med. Chem.*, **26**: 271.

Other systems for generating libraries of polypeptides or nucleotides involve the use of cell-free enzymatic machinery for the *in vitro* synthesis of the library members. In one method, RNA molecules are selected by alternate rounds of selection against a target ligand and PCR amplification (Tuerk and Gold (1990) *Science*, **249**: 505; Ellington and Szostak (1990) *Nature*, **346**: 818). A similar technique may be used to identify DNA sequences which bind a predetermined human transcription factor (Thiesen and Bach (1990) *Nucleic Acids Res.*, **18**: 3203; Beaudry and Joyce (1992) *Science*, **257**: 635; WO92/05258 and WO92/14843). In a similar way, *in vitro* translation can be used to synthesise polypeptides as a method for generating large libraries. These methods which generally comprise stabilised polysome complexes, are described further in WO88/08453, WO90/05785, WO90/07003, WO91/02076, WO91/05058, and WO92/02536. Alternative display systems which are not phage-based, such as those disclosed in WO95/22625 and WO95/11922 (Affymax) use the polysomes to display polypeptides for selection. These and all the foregoing documents also are incorporated herein by reference.

A preferred selection procedure for intracellular immunoglobulins which are stable in an intracellular environment, are correctly folded and are functional with respect to the selective binding of their ligand within that environment is described in WO00/54057. In this approach, the antibody-antigen interaction method uses antigen linked to a DNA-binding domain as a bait and the scFv linked to a transcriptional activation domain as a prey. Specific interaction of the two facilitates transcriptional activation of a selectable reporter gene. An initial in-vitro binding step is performed in which an antigen is assayed for binding to a repertoire of immunoglobulin molecules. Those immunoglobulins which are found to bind to their ligand in vitro assays are then assayed for their ability to bind to a selected antigen in an intracellular environment, generally in a cytoplasmic environment.

INTRACELLULAR CONSENSUS FRAMEWORKS

Intracellular immunoglobulins are advantageously based on an intracellular consensus sequence framework. Advantageously, the consensus is described by at least one of the consensus sequences described in Figure 5 and set forth in Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, T. Y. G. and Rabbitts, T. H. (2002) *J. Mol. Biol.*, **317**, 85-94. Advantageously, the "consensus" used in the present invention is at least 85% identical to that shown in Figure 5; preferably 90%, 95%, 96%, 97%, 98%, 99% or 100% identical thereto. Preferably, in the calculation of identity, the amino acid residues of CDR3 are excluded from consideration.

Intracellular consensus frameworks have been demonstrated, as described in our copending international patent application PCT/GB02/003512, to provide a basis for the construction of immunoglobulins which are stable intracellularly *in vivo*.

The consensus frameworks can be used as a basis for design and construction of immunoglobulins according to the present invention. Oligonucleotides based on the consensus can be used to construct immunoglobulin genes as described; moreover, mutagenesis of CDR sequences using the consensus frameworks allows the generation of libraries of intracellularly active immunoglobulins.

The invention is further illustrated below, in the following examples.

Examples

MATERIALS AND METHODS

Mammalian transactivation domain vector pEF-VP16

The vector pEF-VP16 was constructed for expression of scFv prey in mammalian two hybrid assays. In this vector, scFv sequences may be cloned into Śfi1-Not1 sites in-frame with the VP16 transcriptional transactivator domain (AD) to make a fusion gene controlled by the promoter of the polypeptide elongation factor 1α (EF-1α), which allows high protein expression in mammalian cells (Mizushima, S. and Nagata, S. (1990) *Nucl.*

Acids Res., 18, 5322). The VP16 activation domain fragment, including nuclear localisation signal (nls), was amplified by PCR using pNLVP16 as template and the VP16AD fragment was sub-cloned into the Not1 site of pEF/myc/cyto (Invitrogen). To change the Sfi1 cloning site of pEF/myc/cyto for the Sfi1 site compatible for most scFv fragments, the Sfi1 region of this vector was mutagenised using two oligonucleotides 5'-CGTGAACACGTGGTGGCCCAGCCGGCCCAGGTGCAGC and 5'-GCTGCACCTGGGCCGGCTGGGGGGCCACGTGTTCACG by QuikChange Site-directed Mutagenesis Kit (Stratagene) according to Manufacture's instructions. The final clone has the EF-1α promoter, a multi-cloning site including Sfi1-Not1 sites compatible for scFv fragment insertions, a nuclear localisation signal and the VP16 AD (Figure 1A).

***De novo* antibody gene synthesis**

For antibody gene synthesis, oligonucleotides were designed on an scFv coding sequence comprising the VH and VL framework of the intrabody consensus (Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, T. Y. G. and Rabbitts, T. H. (2002) *J. Mol. Biol.*, **317,** 85-94) and the CDRs of an anti-β-galactosidase scFv R4 (Martineau, P., Jones, P. and Winter, G. (1998) *J Mol Biol*, **280,** 117-127) (Figure 1B). The double strands of DNA were divided into 18 oligonucleotides, of which 16 are 90 bases long and the 2 oligonucleotides flanking the ends of the scFv are respectively 100 bases on the 5'end and 60 bases on the 3'end. Each opposite strand oligonucleotide overlaps by 40-50 bases to ensure good annealing. All crude oligonucleotides were purified on 8% polyacrylamide gels containing 7M urea and visualised by UV shadowing, using fluorescent thin-layer chromatographic plates. Oligonucleotides were eluted by soaking the gel slice in 0.3M sodium acetate overnight at room temperature (~20^{0}C). The supernatant was collected by centrifugation and the oligonucleotides were precipitated with ethanol. The concentration of the purified oligonucleotides were calculated from the absorption spectrum. One μg of each of the purified oligonucleotides was phosphorylated in a final volume of 100μl in the presence of 2μl T4 polynucleotide kinase (10U/μl) and 1mM rATP. The volume was increased to 100μl using NTE (100mM NaCl, 10mM Tris, 1mM EDTA) and phosphorylation carried out by incubation at 37°C for 30 min. The reaction was stopped by incubation at 70°C for 10 min. The oligonucleotides were annealed after boiling the reaction for 30sec and allowing to cool to room temperature (~20^{0}C) over 40 min. Ligation of the annealed oligonucleotides was carried out using 18μl of the annealed

mixture, 2µl of 10X T4 ligase buffer and 1µl T4 DNA ligase (400U/µl) in a final volume of 20µl. The mixture was incubated at 15°C overnight. The assembled oligonucleotides were finally PCR amplified with conseSFI and conseNOT primers (see Figure 1B), which include Sfi1 site at the 5'end and a Not1 site at the 3'end for sub-cloning into pEF-VP16. A master mix for 5 PCR reactions (final volume 30 µl) was prepared containing 500ng of each primer (i.e. conseSFI and conseNOT), 2.5U *pfu* polymerase, 0.2mM dNTPs, 1X PCR reaction buffer and 1µl of the ligated oligo mixture. PCR reaction conditions were denaturation at 94°C for 5 min, followed by 30 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1 min, and a final extension at 72°C for 5 min. The PCR product was separated on a 1% agarose gel and purified using QIAEXII gel purification kit (Quiagen). The purified product (eluted in 40µl elution buffer) and the expression vector was digested with 1µl Sfi1 (10U/µl) in a volume of 30µl at 50°C for 5-6 hours and vector linearisation was checked on an aliquot before proceeding to the Not1 digestion. If the Sfi1 digestions appeared complete, digestion with 1µl Not1 (10U/µl) was carried out at 37°C for 16 hours. The digested PCR products were purified on agarose gels, ligated with vector using T4 ligase at 15°C for 16 hours and transformed into *E.coli* TG-1. The constructs were verified by restriction enzyme digestion using Sfi1 and Not1 and by DNA sequence analysis.

Footprint mutagenesis

Specific mutations of the framework regions of anti-RAS scFv33 (see our copending UK patent application "Anti-activated RAS antibodies" filed on even date herewith) into those of anti-RAS scFvI21 was achieved by PCR-based mutagenesis (herein called footprint mutagenesis). This was done firstly to investigate whether specific amino acid substitutions would affect *in vivo* function of the anti-RAS intrabodies (i.e. antigen binding ability). Anti-RAS scFv33 mutants are listed in Table 1A and were constructed following the flow chart of footprint mutagenesis method shown in Figure 2A. The location of the mutant primer sequences relative to the scFv33 sequence are shown in Figure 2B. Two initial templates were used, either pEF-scFv33-VP16 or pEF-scFvI21-VP16 (respectively scFv33 and scFvI21 cloned in pEF-VP16) for PCR as listed in Table 1. Each mutagenesis comprised synthesis of two overlapping PCR products using mutant oligonucleotides (Figure 2A, step 1) followed by complete assembly (Figure 2A, step 2) and cloning into the pEF-VP16 vector (Figure 2A, step 3) to generate the mutated

template for the next round of footprint mutagenesis (repeat). At each step the functional validity of the changes was estimated. Step 1 PCR reactions (final volume 20 µl) contained 0.5 µM of each primer pair, 2.5U *pfu* polymerase, 0.2mM dNTPs, 1X PCR reaction buffer and 50 ng of pEF-scFv-VP16 template. PCR reactions were carried out by denaturation at 95°C for 5min, followed by 30 cycles of 95°C for 30sec, 60°C for 30sec and 75°C for 45sec, and a final extension at 75°C for 10min. Following PCR amplification, the amplified DNA fragments were electrophoresed on 2% agarose, extracted and purified by using QIAquick Gel Extraction Kit (Qiagen). Purified PCR fragments were assembled and amplified by PCR in Step 2 with pEF-VP16 vector primers EFFP (5'-TCTCAAGCCTCAGACAGTGGTTC-3') and VP162R (5'-CAACATGTCCAGATCGAA -3') by denaturation at 95°C for 5min followed by a gradient annealing at 60°C to 30°C (0.1°C per sec reduction in temperature), and gradient extension at 30°C to 75°C (0.1°C per sec, increasing temperature) followed by 29 cycle with denaturation at 95°C for 45sec, annealing at 60°C for 45sec and extension at 75°C for 90sec. The amplified DNA fragment was digested with Sfi1 and Not1, purified by electrophoresis and gel extraction and re-cloned into Sfi1 and Not1 site of pEF-VP16 in Step 3. The constructs were verified by restriction enzyme digestion using Sfi1 and Not1 and confirmed by DNA sequencing and tested for antigen binding *in vivo*.

The construction of pEF-scFvI21R33-VP16 (i.e. scFvI21R33 has CDRs of anti-RAS scFv 33 and frameworks of anti-RAS scFvI21 except the lysine at VH position 94 was changed to arginine) was performed by repeated the PCR-assembly-cloning procedures described above according to Table 1B, starting with Mut1 as a template. Each round of mutation gave the mutated template for the next round of the stepwise footprint mutagenesis using conditions described above.

Diversification of VH CDR3 by footprint mutagenesis for intrabody library construction

A flow chart outlining the construction of an scFv library based on VH CDR3 randomisation is shown in Figure 3A. The primers used to randomise CDR3 of the VH domain in anti-ABL scFvA25 (Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, T. Y. G. and Rabbitts, T. H. (2002) *J. Mol. Biol.*, **317,** 85-94) and partial nucleotide and protein sequence of anti-ABL scFvA25 are shown in Fig 3B. The CDR3 randomisation was performed using footprint mutagenesis. The template encoding anti-ABL scFvA25

was sub-cloned into the pEF-VP16 vector. In step 1, two PCR products were made using the pEF-scFvA25-VP16 template viz. the VH domain plus JH using the PCR primers EFFP2 plus A25C3B$_n$ and the VL using the PCR primers A25CDR3F plus VP162R. The two PCR reactions yield overlapping products (Figure 3A). The A25C3B$_n$ comprised three distinct oliognucleotides, each with an homologous sequence footprint around mutagenic regions of 3, 6 and 10 codons to generate mutations within VH CDR3. Amplified PCR fragments (for VH and VL regions) were individually electrophoresed on agarose and purified. The two PCR products were assembled in a second PCR reaction using oligonucleotides EFFP andVP162R which encompass the whole scFv (i.e. VH and VL). The final PCR product was digested with Sfi1 and Not1 and ligated with Sfi1-Not1-digested pEF-VP16. Ligated DNA were electroporated in the E.coli DH10ß (Invitrogen). Clones were randomly picked from each final ligation (i.e. from A25C3B3, A25C3B6 and A25C3B10) and sequenced to verify the insert and the correct integration of CDRs.

Primer sequences:-

EFFP2: 5'- GGAGGGGTTTTATGCGATGG-3'.

EFFP: 5'-TCTCAAGCCTCAGACAGTGGTTC-3'

A25C3B: 5'-GACGGTGACCAGGGTTCCCTGGCCCC(A/CNN)$_N$TCTCGCACAGTATATTAC-3', where n=3, 6 or 10 to randomise amino acid residues in CDR3 of VH domain.

A25CDR3F: 5'-GGGGCCAGGGAACCCTGGTCACCGTC-3'.

VP162R: 5'- CAACATGTCCAGATCGAA-3'.

Diversification of VH CDRs by footprint mutagenesis for intrabody library construction

A flow chart outlining the randomisation of the VH CDRs is shown in Figure 4A. Two templates were used. One encoding the VH domain from anti-RAS scFvI21R33 and the other from the canonical intrabody consensus sequence (Tse *et al.*, *supra*), each sub-cloned into the pEF-VP16 vector.

Library 1 (CDR2/3): Randomisation of VH CDR 2 and 3 in each case was done by footprint mutagenesis as described above. In the first round of PCR amplification (Fig. 4A, step 1), two parts of the VH domain were separately amplified by PCR using two pairs of oligonucleotides: EFFP2 plus CDR2R (to randomise CDR2) and CDR2F plus CDR3R (to randomise CDR3). Amplified PCR fragments were electrophoresed on

agarose and purified. In the second round of PCR (Fig. 4A, step 2), the two PCR fragments were assembled using PCR oligonucleotides EFFP2 and JH5R. After purification of PCR product, a final PCR (Fig. 4A, step 3) was performed using EFFP and NotVHJR1 to allow digestion with Sfi1 and Not1 and ligation into yeast pVP16* vector cut with Sfi1 + Not1. Ligated DNA was electroporated in competent E.coli DH10B. This facilitated the generation of two libraries (each called VH CDR2 /3 library 1) with diversities of 2×10^6 (I21R33-derived library) and 1.4×10^6 (consensus library)

Library 2 (CDR1/2/3): For randomisation of VH CDR1, the two CDR2/3 libraries (library 1) were used as templates. Two PCR reactions were carried out with pair of oligonucleotides: sFvVP16F plus CDR1R (to randomise CDR1) and CDR1F and VP162R (to copy the remaining part of the VH segment) (Figure 4B). The two PCR fragments were assembled using sFvVP16F and VP162R, digested with Sfi1 and Not1 and ligated into yeast pVP16* vector cut with Sfi1 and Not1. This facilitated the generation of two libraries size of library 2 with diveristies of 3.04×10^7 (I21R33-derived library) and 2.215×10^7 (consensus library). Clones were randomly picked up from each library and sequenced to verify the insert and the correct integration of CDRs (Fig. 4C).

Primer sequences:-

CDR2R: 5'-CAGAGTCTGCATAGTATAT$(MNN)_5$ACTAATGTATGAAACCCAC-3'

CDR2F: 5'- ATATACTATGCAGACTCTG -3'

CDR3R: 5'-TCCCTGGCCCCAGTAGTCAAA(MNNMNN)nCCCTCTCGCACAGTAATAG-3', where n=1 to 6 to randomise amino acid residues in CDR3 of VH domain.

JH5R, 5'-GGTGACCAGGGTTCCCTGGCCCCAGTAGTC-3'

NotVHJR1; 5'-ATAAGAATGCGGCCGCCGCTCGAGACGGTGACCAGGGTTCCCTG-3'.

sFvVP16F, 5'- TGGGTCCGCCAGGCTCCAGG -3'.

CDR1R: 5'-CCTGGAGCCTGGCGGACCCAMNNCAT$(MNN)_3$CTGAAGCTGAATCCAGAGG-3'

CDR1F, 5'-TGGGTCCGCCAGGCTCCAGG-3'

Mammalian two hybrid assay in CHO-CD4 using FACS analysis

Chinese hamster ovary (CHO) cells were grown in minimal essential medium α (α-MEM, Invitrogen) with 10% foetal calf serum, penicillin and streptomycin. FACS analysis using the CHO-CD4 reporter line (Fearon, E. R., Finkel, T., Gillison, M. L., Kennedy, S. P., Casella, J. F., Tomaselli, G. F., Morrow, J. S. and Dang, C. V. (1992) *Proc. Natl. Acad. Sci. USA*, **89,** 7958-7962) was performed as described previously (Tse *et al., supra*) with small modifications. 3 x 10^5 CHO-CD4 cells were seeded in 6 well plates on the day before transfection. 0.5 μg of pM1-HRASG12V (DBD-RAS) or pM1-ßgal (DBD-βgal) and 1 μg of pEF- scFvR4-VP16 or pEF-scFvconR4-VP16 were co-transfected into the cells using lipofectAMINE™ according to Manufacture's instructions. Forty-eight hours after transfection, cells were washed, dissociated using cell dissociation solution (Sigma) and re-suspended in PBS. The induction of cell surface CD4 expression was detected using anti-human CD4 antibody (Pharmingen) and FITC-conjugated anti-mouse IgG (Pharmingen). The relative fluorescence of the cells were measured with a FACSCalibur (Becton Dickinson) and the data were processed using the CELLQuest software.

RESULTS AND DISCUSSION

***De novo* antibody gene synthesis**

The production of antibody V-genes from known protein sequence data was carried out by development of a set of overlapping oligonucleotides corresponding to the intracellular antibody scFv consensus framework (Tse *et al., supra*) together with VH and VL CDRs from the anti-βgal scFv R4 (18) (Figure 1B). Annealing and ligation of the mixture of oligonucleotides was followed by PCR of the assembled scFv and finally cloning into the mammalian expression vector pEF-VP16 vector after Sfi1-Not1 digestion (Figure 1A; pEF-VP16). The synthetic scFv was cloned to derive pEF-scFvconR4-VP16 and this was sequenced to verify the scFv and its junction with the VP16 AD domain. The effectiveness of the hybrid scFvconR4 as an intrabody was assayed in a reporter assay, co-transfecting the pEF-scFvconR4-VP16 clone plus DBD-lacZ (Tse, E. and Rabbitts, T. H. (2000) *Proc. Natl. Acad. Sci. USA*, **97,** 12266-12271) into the CHO-CD4 line (this line carries a *CD4* gene with a minimal promoter regulated by five repeated *Gal4* DNA-binding sites and can be transcriptionally activated and monitored by expression of cell surface CD4 (Fearon, E. R., Finkel, T., Gillison, M. L., Kennedy, S. P., Casella, J. F.,

Tomaselli, G. F., Morrow, J. S. and Dang, C. V. (1992) *Proc. Natl. Acad. Sci. USA*, **89,** 7958-7962)). When pEF-scFvconR4-VP16 was transfected into CHO-CD4 with a clone encoding a *Gal4* DNA binding domain (DBD) fused to βgal, we detected around activation of CD4 expression (Figure 1C). This compares with analogous experiments using the DBD-βgal with the original scFvR4 (as pEF-scFvR4-VP16). However, no activation of CD4 expression was observed with either pEF-scFvconR4-VP16 or pEF-scFvR4-VP16 co-transfected with a non-relevant bait, DBD-RAS. This shows that the *de novo* gene synthesis method is efficient for cloning antibody fragments which retain their specificity and verifies the consensus framework as an intrabody expression scaffold.

Footprint mutagenesis to create intrabody diversity

The intracellular antibody capture method defined an scFv consensus sequence which proved particularly advantageous for intracellular use (Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, T. Y. G. and Rabbitts, T. H. (2002) *J. Mol. Biol.*, **317,** 85-94) because the method selects intrabodies based on *in vivo* screens (Visintin, M., Tse, E., Axelson, H., Rabbitts, T. H. and Cattaneo, A. (1999) *Proc. Natl. Acad. Sci. USA*, **96,** 11723-11728; Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, T. Y. G. and Rabbitts, T. H. (2002) *J. Mol. Biol.*, **317,** 85-94; Visintin, M., Settanni, G., Maritan, A., Graziosi, S., Marks, J. D. and Cattaneo, A. (2002) *J. Mol. Biol.*, **317,** 73-83). One specific antibody derived using this method was scFv33, which is an anti-RAS antibody able to bind RAS in mammalian cells. A second scFv, scFvI21, was derived from a RAS yeast screen but did not bind RAS in mammalian cells, although its expression level was superior to scFv33. We wished to assess the importance of specific IAC consensus framework residues and we have used a PCR-based mutagenesis procedure (herein called footprint mutagenesis) to make mutations in the scFv33 framework (Table 1A) for evaluation in a mammalian cell luciferase reporter assay (Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, T. Y. G. and Rabbitts, T. H. (2002) *J. Mol. Biol.*, **317,** 85-94). With a series of changes, scFv33 framework was effectively converted to scFvI21 in a step-wise manner (Table 1) which exemplifies the consensus framework and which retains parental ability to bind to RAS antigen (summarised in Table 1A). The templates scFv33 and scFvI21 were cloned into pEF-VP16 to be used as PCR templates and sequential mutagenesis was carried out. Full conversion required seven rounds of PCR, assembly and cloning (Table 1B; note the only change to the scFvI21 framework was the lysine residue at position 94

was changed to an arginine, consistent with the canonical intrabody consensus). At each step a new template was created and sequenced to verify the specificity of the PCR and each mutation was tested for function. For sequential mutation, each round provides the template for next mutagenesis step (Table 1B; Figure 2). With this approach, a new scFv could be created with the framework of scFvI21R and the CDRs of the scFv33 (scFvI21R33).

Using footprint mutagenesis to diversify CDRs and create intrabody libraries

Footprint mutagenesis can be applied to create single or small changes in specified region. We have used this method to diversify one CDR in an scFv (VH CDR3 of the anti-ABL scFvA25 (Tse, E., Lobato, M. N., Forster, A., Tanaka, T., Chung, T. Y. G. and Rabbitts, T. H. (2002) *J. Mol. Biol.*, **317**, 85-94)) to generate a library of different sequences. Footprint mutagenesis was achieved in this case using as the template scFvA25 cloned in pEF-VP16 and an internal mixture of PCR primers covering scFv VH CDR3 in which the oligonucleotides contain three, six or ten codon equivalence of randomised sequence (Figure 3A, B; primer A25C3B$_n$). After the first PCR, the two PCR products were assembled (Figure 3A, 2nd PCR) and the final product was cloned into pEF-VP16 to create a library of individual clones. Randomly picked clones were sequenced in the CDR3 region (Figure 3C) confirming that the CDR3 has been changed by 3, 6 or 10 residues respectively. Thus this simple footprint mutagenesis method generates high degrees of diversity which is directly related to design of the primers used for the PCR steps.

In the procedure illustrated in Figures 2 and 3, each step introduced mutation(s) at only one position in the scFv. Mutations can be made at two, and potentially more, positions by using mutant oligonucleotides for each PCR step, prior to assembly. This is shown in Figure 4 which illustrates simultaneously randomising the CDR2 and CDR3 regions of a VH template and subsequent randomisation of CDR1 (the same strategy would apply to mutagenesis of VL). In the examples for CDR2/3 changes, two VH sub-region PCR reactions were carried out, each using a fixed sequence primer together with a randomising primer (Figure 4A, EFFP2 + CDR2R or CDRF + CDR3R, where the two reverse primers are mutagenic for CDR2 and CDR3 respectively). The two PCR products overlap in the CDR2 region and a PCR assembly was achieved using the flanking primers, followed by final PCR with EFFP plus NOTVHJR1 for cloning into the Sfi1-

Not1 site of yeast pVP16* vector. Sequences of a selection of clones showed diversity of the CDR2 and CDR 3 regions (Figure 4C). A mixed library of 3.4 X 10^6 clones from the above was used as a substrate for a second round of mutagenesis at CDR1 using the pairs of primers EFFP2 plus CDR1R (a primer for mutagenic CDR1) and CDR1F plus VP162R. Sequences of a selection of clones showed diversity of the CDR1, as well as at CDR2 and CDR 3 regions (Figure 4C). This two step procedure thus allows production of randomised CDR1, 2 and 3. It should be possible to devise a similar protocol for simultaneously mutating CDR1, 2 and 3.

In summary, the *de novo* antibody gene synthesis method and footprint mutagenesis are powerful tools for making antibody genes and the acquisition of immunoglobulin mutants, for instance for affinity maturation which would involve CDR changes. We show that the *de novo* antibody gene synthesis method is a simple, oligonucleotide-based annealing, ligation and PCR procedure to make an antibody fragment suitable for cloning into a compatible vector. In our specific use of *de novo* intrabody production, a hybrid scFv was made in which the intracellular antibody capture (IAC) consensus was the scaffold and an anti-βgal antibody (Martineau, P., Jones, P. and Winter, G. (1998) *J Mol Biol*, **280,** 117-127) provided the CDR sequences. We chose the IAC consensus because it is advantageous for mammalian in-cell expression and anti-βgal antibody because it was specially developed in bacteria for soluble expression. Our hybrid scFv was able to bind to its target antigen in CHO cells with comparable efficiency to the parental scFv (Figure 1C). This adds further validation to the IAC consensus scaffold as a suitable intrabody scaffold and shows that *de novo* intrabody gene production is viable. Diversification of the antibody fragments was carried out by footprint mutagenesis allowing one or two step conversion of a V-gene. Generation of whole intrabody libraries was achieved by this means. Thus intrabody libraries can be made starting with the consensus intrabody sequence and using these two simple *in vitro* methods. These libraries are ready for direct *in vivo* screening with any antigen that can be made as a bait in a suitable yeast two-hybrid vector.

A

anti-RAS scFv33 mutants	template	Forward Primer	Reverse Primer	in vivo function
VH(Q1E+V5L+A7S+S28T) **(Mut1)**	I21*	EFFP2	SFVI21R	+
	33*	SFV33F	VP162R	
VH(A74S+S77T)	33	EFFP2	33MUT1R	+
	33	33MUT1F	VP162R	
VH(D84A)	33	EFFP2	33MUT2R	+
	33	33MUT2F	VP162R	
VH(R94K)	33	EFFP2	33MUT6R	-
	33	33MUT6F	VP162R	
VL(0T+V3Q)	33	EFFP2	33MUT3R	+
	33	33MUT3F	VP162R	
VL(F10S)	33	EFFP2	33MUT4R	+
	33	33MUT4F	VP162R	
VL(I84T)	33	EFFP2	33MUT5R	+
	33	33MUT5F	VP162R	
VL(G100Q+V104L)	33	EFFP2	SFV33R	+
	I21	SFVI21F	VP162R	

I21*, pEF-scFvI21-VP16; 33*, pEF-scFv33-VP16

B

anti-RAS scFv33 mutants	template*	Forward Primer	Reverse Primer
VH(Q1E+V5L+A7S+S28T)+VL(G100Q+V104L) **(Mut2)**	Mut1	EFFP2	SFV33R
	I21	SFVI21F	VP162R
Mut2+VH(A74S+S77T) **(Mut3)**	Mut2	EFFP2	33MUT1R
		33MUT1F	VP162R
Mut3+VH(D84A) **(Mut4)**	Mut3	EFFP2	33MUT2R
		33MUT2F	VP162R
Mut4+VL(0T+V3Q) **(Mut5)**	Mut4	EFFP2	33MUT3R
		33MUT3F	VP162R
Mut5+VL(F10S) **(Mut6)**	Mut5	EFFP2	33MUT4R
		33MUT4F	VP162R
I21R33	Mut6	EFFP2	33MUT5R
		33MUT5F	VP162R

template*, pEF-scFv-VP16

Table 1. Templates and primers for stepwise footprint mutagenesis to convert scFv33 to scFvI21R33

The primers were used in footprint mutagenesis, illustrated in Figure 2, of the framework of scFv33 sequence, to convert it to the I21R33 sequence.

A. At the first round, both pEF-scFv33-VP16 and pEF-scFvI21-VP16 were used as templates. Individual mutations were incorporated with the primers, as indicated.

B. At subsequent rounds, the PCR template used was the previously mutated version, except round 2 in which either Mut1 or pEF-scFvI21R-VP16 were used.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are apparent to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Claims

1. A method for preparing a double stranded nucleic acid which encodes an immunoglobulin, comprising the steps of:

(a) providing a set of three or more overlapping oligonucleotides which anneal to form the + and – strands of a nucleic acid which encodes at least part of an immunoglobulin variable domain;

(b) annealing the oligonucleotides;

(c) replicating the + and – strands of the nucleic acid formed from the annealed oligonucleotides; and

(d) inserting the nucleic acid into an expression vector.

2. A method according to claim 1, wherein the double stranded nucleic acid encodes an immunoglobulin variable domain.

3. A method according to claim 1 or claim 2, wherein the variable domain is a V_H domain.

4. A method according to claim 1 or claim 2, wherein the variable domain is a V_L domain.

5. A method according to claim 1 or claim 2, wherein the immunoglobulin is a scFv and the nucleic acid encodes both a V_H domain and a V_L domain, linked by a linker sequence.

6. A method according to any preceding claim, wherein the nucleic acid further encodes at least one effector domain or a label.

7. A method according to any preceding claim, comprising the further steps of:

(e) amplifying the nucleic acid encoding the immunoglobulin variable domain using a first set of at least four primers, of which two primers are overlapping, and wherein one of the overlapping primers consists of a plurality of different primers, each

of which comprises a mutagenic region which generates a mutation in a part of the nucleic acid sequence;

(f) purifying the amplification products thus obtained;

(g) assembling the amplification products in a further amplification reaction using a second set of primers which encompass the entire nucleic acid which encodes the immunoglobulin variable domain; and

(h) inserting the assembled amplification product into an expression vector.

8. A method according to claim 7 wherein the nucleic acid and assembled amplification product which is finally obtained after the mutagenesis procedure encode the gene product which it is desired to obtain in its entirety.

9. A method according to claim 7 or claim 8, wherein primers are used which comprise a multiple codon equivalence of randomised or semi-randomised sequence.

10. A method according to any one of claims 7 to 9, wherein steps (e) to (g) are repeated, using different primers, to mutate different regions of the nucleic acid and thus generate a library of diversified nucleic acid molecules which encompass mutations in a plurality of regions.

11. A method according to any one of claims 7 to 9, wherein two or more diversified primer sets are used simultaneously to diversify two or more separate regions.

12. A method according to any one of claims 7 to 11 wherein two or more PCR products are generated which overlap.

13. A method according to any one of claims 7 to 12, wherein multiple codon equivalence extends over two to twelve codons.

14. A method for preparing a library of nucleic acids encoding a diversified immunoglobulin, comprising the steps of:

(a) amplifying a nucleic acid encoding the immunoglobulin using a first set of at least four primers, of which two primers are overlapping, and wherein one of the

overlapping primers consists of a plurality of different primers, each of which comprises a mutagenic region which generates a mutation in a part of the nucleic acid;

(b) purifying the amplification products thus obtained;

(c) assembling the amplification products in a further amplification reaction using a second set of primers which encompass the entire nucleic acid which encodes the immunoglobulin; and

(d) inserting the assembled amplification product into an expression vector.

15. A method according to claim 14 wherein the nucleic acid and assembled amplification product which is finally obtained after the mutagenesis procedure encode the gene product which it is desired to obtain in its entirety.

16. A method according to claim 14 or claim 15, wherein primers are used which comprise a multiple codon equivalence of randomised or semi-randomised sequence.

17. A method according to any one of claims 14 to 16, wherein two or more diversified primer sets are used simultaneously to diversify two or more separate regions.

18. A method according to any one of claims 14 to 17 wherein two or more PCR products are generated which overlap.

19. A method according to any one of claims 14 to 18, wherein multiple codon equivalence extends over two to twelve codons.

20. The method of any preceding claim, further comprising the steps of causing the expression vector to produce the nucleic acid molecule and isolating the immunoglobulin thus produced.

21. An immunoglobulin produced by expression of an expression vector according to any one of claims 1 to 19.

22. A library produced by the method of any one of claims 14 to 19.

FIG. 1

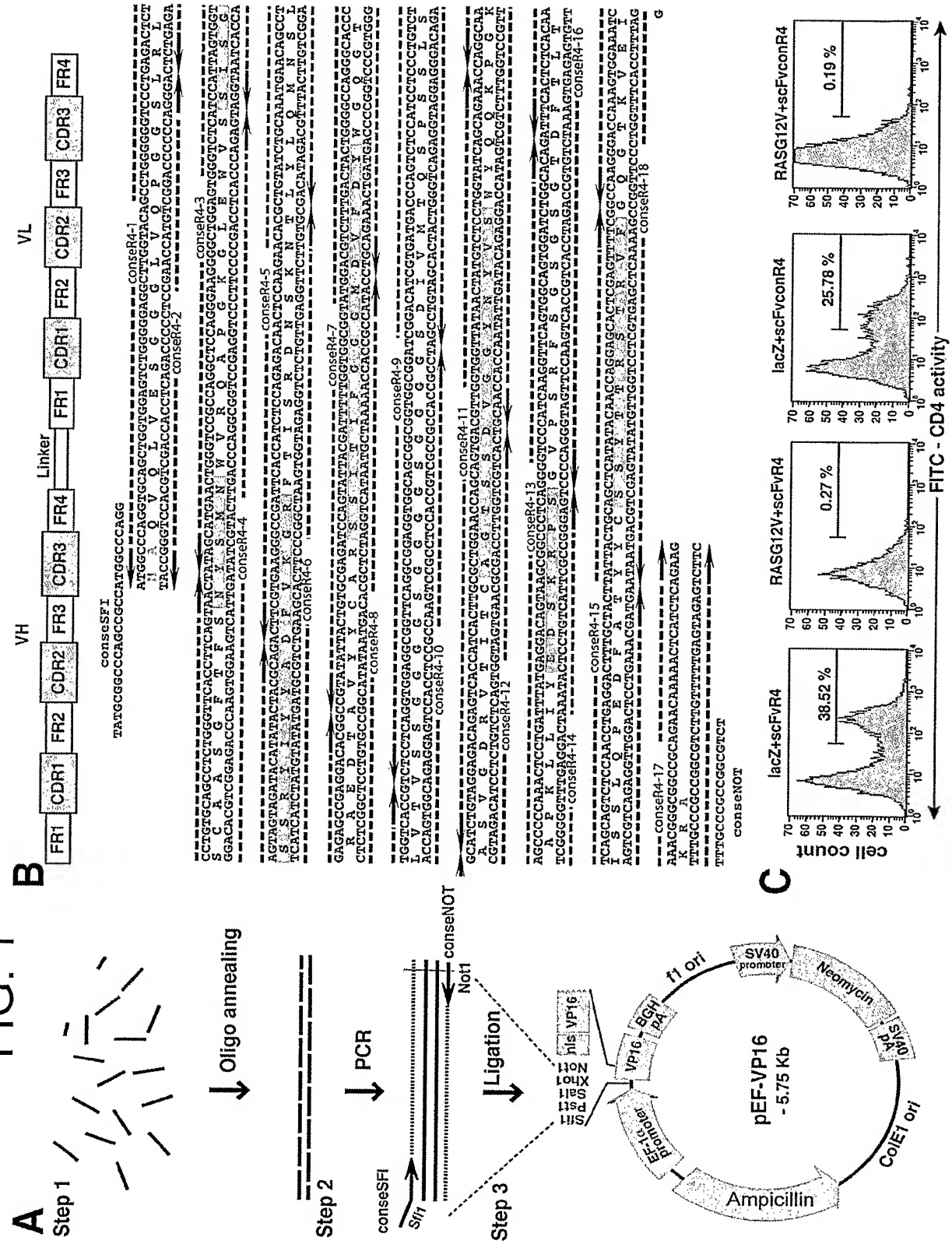

A
Step 1
Oligo annealing
Step 2
PCR
conseSFI
SfiI
conseNOT
Not1
Step 3
Ligation
pEF-VP16
- 5.75 Kb
Ampicillin
ColE1 ori
Ef-1alpha Promoter
VP16
BGH pA
f1 ori
SV40 promoter
Neomycin
SV40 pA
B
VH
VL
Linker
FR1
CDR1
FR2
CDR2
FR3
CDR3
FR4
conseSFI
conseR4-1
conseR4-2
conseR4-3
conseR4-4
conseR4-5
conseR4-6
conseR4-7
conseR4-8
conseR4-9
conseR4-10
conseR4-11
conseR4-12
conseR4-13
conseR4-14
conseR4-15
conseR4-16
conseR4-17
conseR4-18
conseNOT
C
lacZ+scFvR4
38.52 %
RASG12V+scFvR4
0.27 %
lacZ+scFvconR4
25.78 %
RASG12V+scFvconR4
0.19 %
cell count
FITC - CD4 activity

FIG. 2

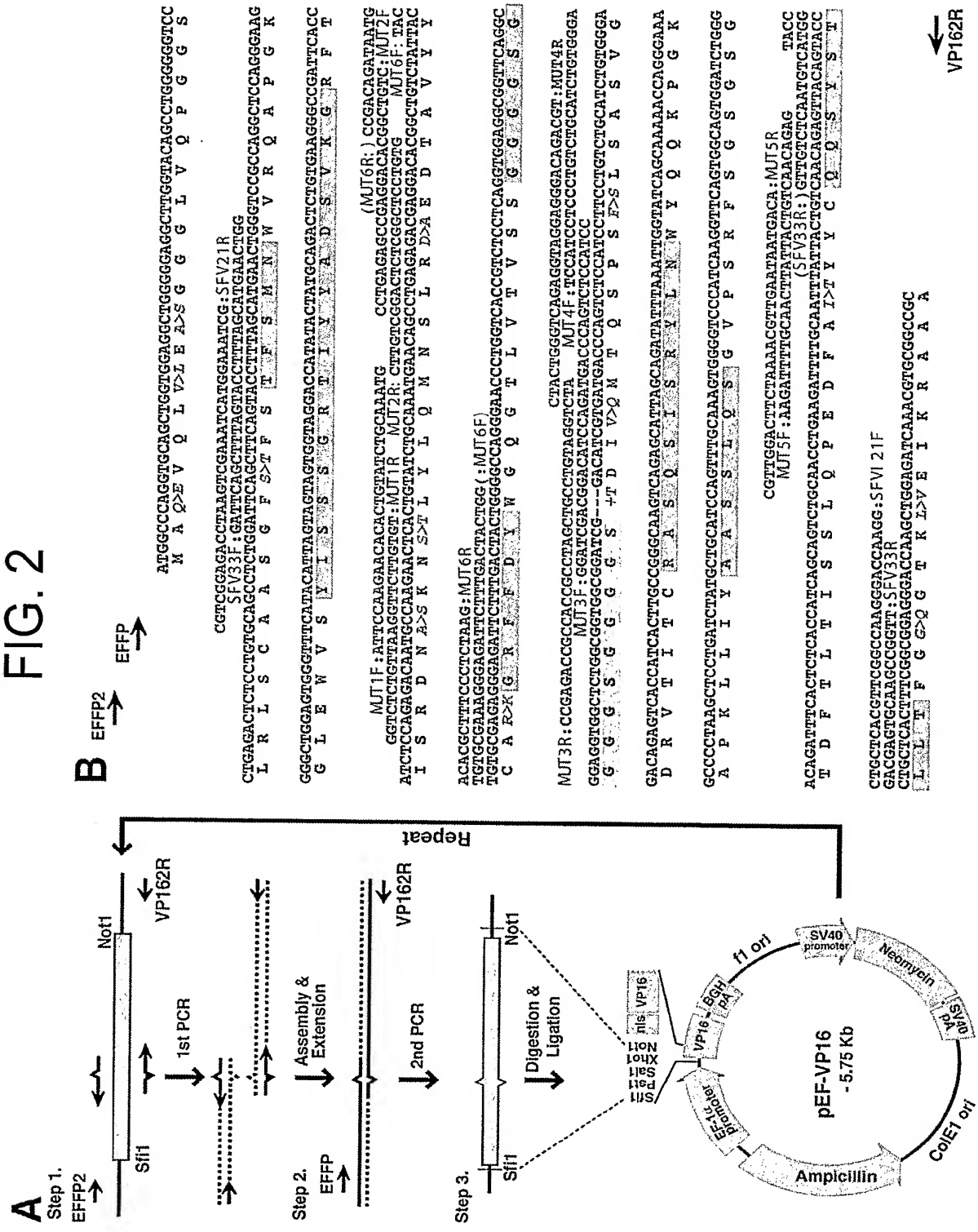

A
Step 1.
EFFP2
Sfi1
Not1
VP162R
1st PCR
Step 2.
Assembly & Extension
EFFP
2nd PCR
VP162R
Step 3.
Sfi1
Not1
Digestion & Ligation
Repeat
pEF-VP16
- 5.75 Kb
Ampicillin
ColE1 ori
EF-1α promoter
VP16
BGH pA
f1 ori
SV40 promoter
Neomycin
SV40 pA
nls VP16
Sfi1
Pst1
Sal1
Xho1
Not1
B
EFFP2
EFFP
ATGGCCCAGGTGCAGCTGGTGGAGGCTGGGGGAGGCTTGGTACAGCCTGGGGGGTCC
M A Q>E V Q L V>L E E A>S G G G G L V Q P G G G S
CGTCGGAGACCTAAGTCGAAATCATGGAAATCG:SFV21R
SFV33F:GATTCAGCTTTAGTACCTTTAGCATGAACTGG
CTGAGACTCTCCTGTGCAGCCTCTGGATTCAGCTTTAGCATGAACTGGGTCCGCCAGGCTCCAGGGAAG
L R L S C A A S G G F S>T F S M N W V R Q A P G K
GGGCTGGAGTGGGTTTCATACATTAGTAGTGGTAGGACCATATACTATGCAGACTCTGTGAAGGGCCGATTCACC
G L E W V S Y I S S G R T I Y Y A D S V K G R F T
MUT1F:ATTCCAAGAACACACTGTATCTGCAAATG CCTGAGAGCCGAGGACACGGCTGTC:MUT2F
GGTCTCTGTTAAGGTTCTTGTGT:MUT1R MUT2R:CTTGTCGGACTCTCGGCTCCTGTG MUT6F:TAC
ATCTCCAGAGACAATGCCAAGAACTCACTGTATCTGCAAATGAACAGCCTGAGAGACGAGGACACGGCTGTCTATTAC
I S R D N A>S K N S>T L Y L Q M N S L R D>A E D T A V Y Y
(MUT6R:) CCGACAGATAATG
ACACGCTTTCCCTCTAAG:MUT6R
TGTGCGAAAGGGAGATTCTTTGACTACTGG(:MUT6F)
TGTGCGAGAGGGCGATTCTTTGACTACTGGGGCCAGGGAACCCTGGTCACCGTCTCCTCAGGTGGAGGCGGTTCAGGC
C A R>K G R F F D Y W G Q G T L V T V S S G G G G S G
CTACTGGGTCAGAGGTAGGAGGGACAGACGT:MUT4R
MUT3R:CCGAGACCGCCACCGCCTAGCTGCCTGTAGGTCTA MUT4F:TCCATCCTCCCTGTCTGCATCTGTGGGA
MUT3F:GGATCGACGGACATCCAGATGACCCAGTCCATCC
GGAGGTGGCTCTGGCGGTGGCGGATCG---GACATCGTGATGACCCAGTCTCCATCCTCCCTGTCTGCATCTGTGGGA
G G G S G G G G S +T D I V>Q M T Q S P S F>S L S A S V G
DRVTITC RASQSISRYLNWYQQKPGK
GACAGAGTCACCATCACTTGCCGGGCAAGTCAGAGCATTAGCAGATATTTAAATTGGTATCAGCAAAAACCAGGGAAA
GCCCCTAAGCTCCTGATCTATGCTGCATCCAGTTTGCAAAGTGGGGTCCCATCAAGGTTCAGTGGCAGTGGATCTGGG
APKLLIYAASSLQSGVPSRFSGSGSG
CGTTGGACTTCTAAAACGTTGAATTAATGACA:MUT5R
MUT5F:AAGATTTTGCAACTTATTACTGTCAACAGAG TACC
(SFV33R:)GTTGTCTCAATGTCATGG
ACAGATTTCACTCTCACCATCAGCAGTCTGCAACCTGAAGATTTTGCAATTTATTACTGTCAACAGAGTTACAGTACC
T D F T L T I S S L Q P E D F A I>T Y Y C Q Q S Y S T
VP162R
CTGCTCACGTTCGGCCAAGGGACCAAGG:SFVI 21F
GACGAGTGCAAGCCGGTT:SFV33R
CTGCTCACTTTGGCGGAGGACCAAGCTGGAGATCAAACGTGCGGCCGC
L L T F G G>Q G T K L>V E I K R A A A

FIG. 3

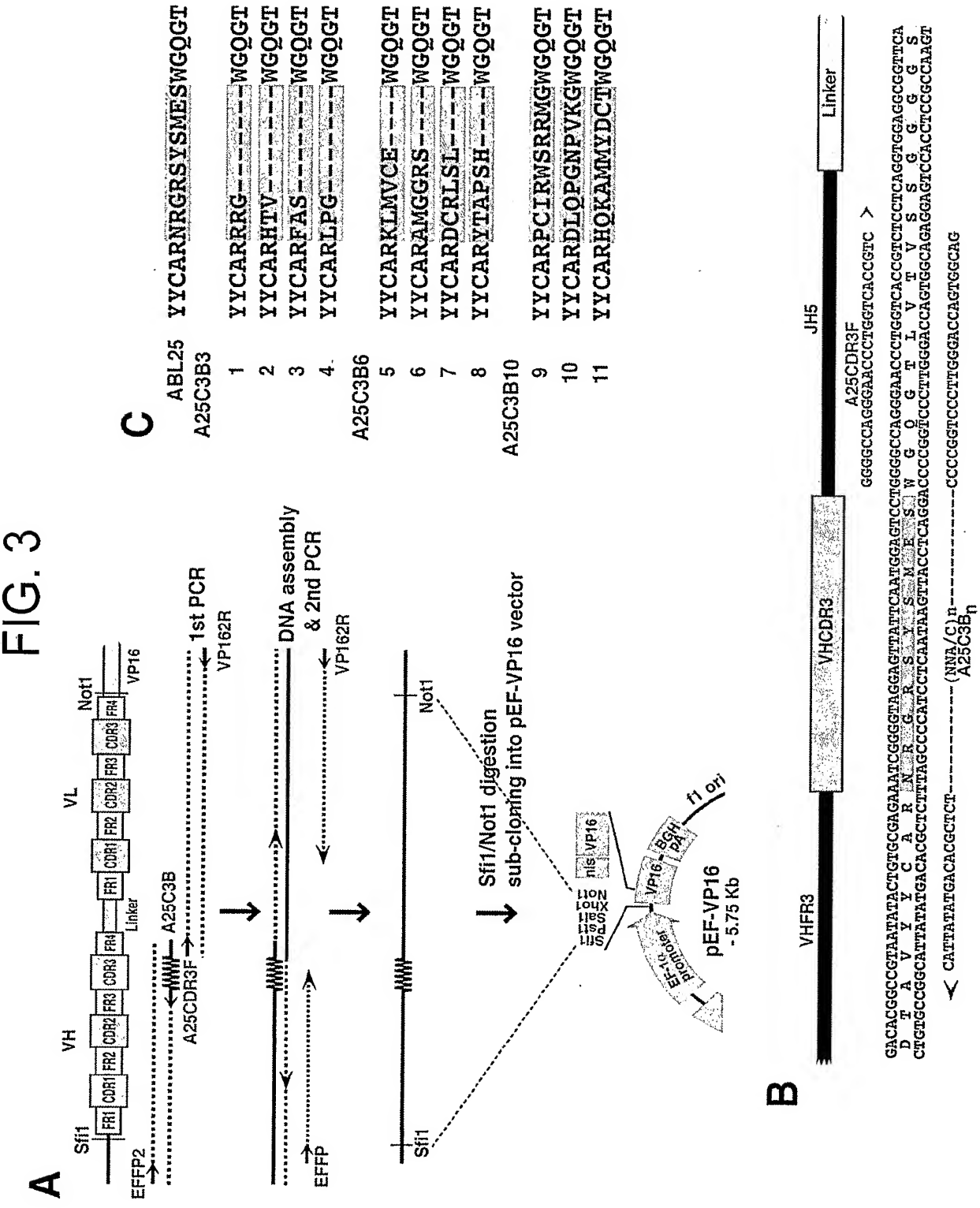
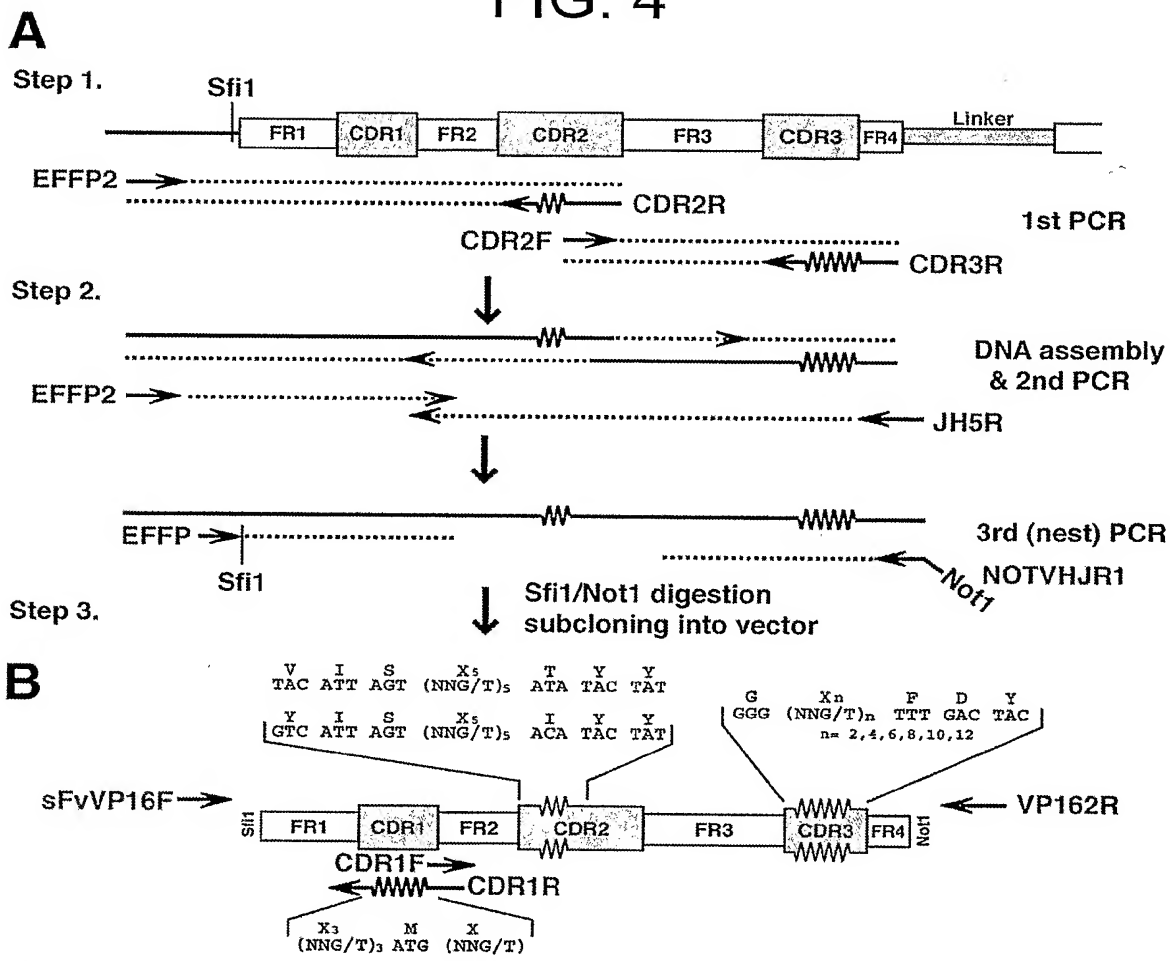

A
VH
VL
Sfi1
Not1
FR1
CDR1
FR2
CDR2
FR3
CDR3
FR4
Linker
VP16
EFFP2
A25C3B
A25CDR3F
1st PCR
VP162R
DNA assembly
& 2nd PCR
EFFP
Sfi1/Not1 digestion
sub-cloning into pEF-VP16 vector
pEF-VP16
- 5.75 Kb
f1 ori
BGH pA
EF-1α promoter
B
VHFR3
VHCDR3
JH5
Linker
A25CDR3F
C
ABL25 YYCARNRGRSYSMESWGQGT
A25C3B3
1 YYCARRRG----------WGQGT
2 YYCARHTV----------WGQGT
3 YYCARFAS----------WGQGT
4 YYCARLPG----------WGQGT
A25C3B6
5 YYCARKLMVCE-------WGQGT
6 YYCARAMGGRS-------WGQGT
7 YYCARDCRLSL-------WGQGT
8 YYCARYTAPSH-------WGQGT
A25C3B10
9 YYCARPCIRWSRRMGWGQGT
10 YYCARDLQPGNPVKGWGQGT
11 YYCARHQKAMMYDCTWGQGT

FIG. 4



C

		CDR1	CDR2	CDR3
Con		SYAMH	VISGDGSNTYYADSVKG	G.....DY
PCR oligo		XXXMX	VISXXXXXTYYADSVKG	GXXXXXYXXXXXXFDY
CDR2/3	1	TFSMN	YISADERIIYYADSVKG	GPP----------FDY
	2	TFSMN	YISYGALCIYYADSVKG	GGWHS---------FDY
	3	TFSMN	YISASNRAIYYADSVKG	GIPFGLINLRLLMFDY
	4	SYAMH	VISTHITQTYYADSVKG	GCYRK---------FDY
	5	SYAMH	VISRIIKDTYYADSVKG	GCNGNYHRMAG--FDY
CDR1/2/3	1	DRNMD	YISSVTTVIYYADSVKG	GYLKHRRCGIATIFDY
	2	TGPML	YISSSPTNIYYADSVKG	GEAMNYT------FDY
	3	TNRMQ	VISTTKAYTYYADSVKG	GLL----------FDY
	4	RVPMS	VISAKIQKTYYADSVKG	GGICW---------FDY
	5	PRQML	VISVRHYETYYADSVKG	GGRRE---------FDY

FIG. 5A

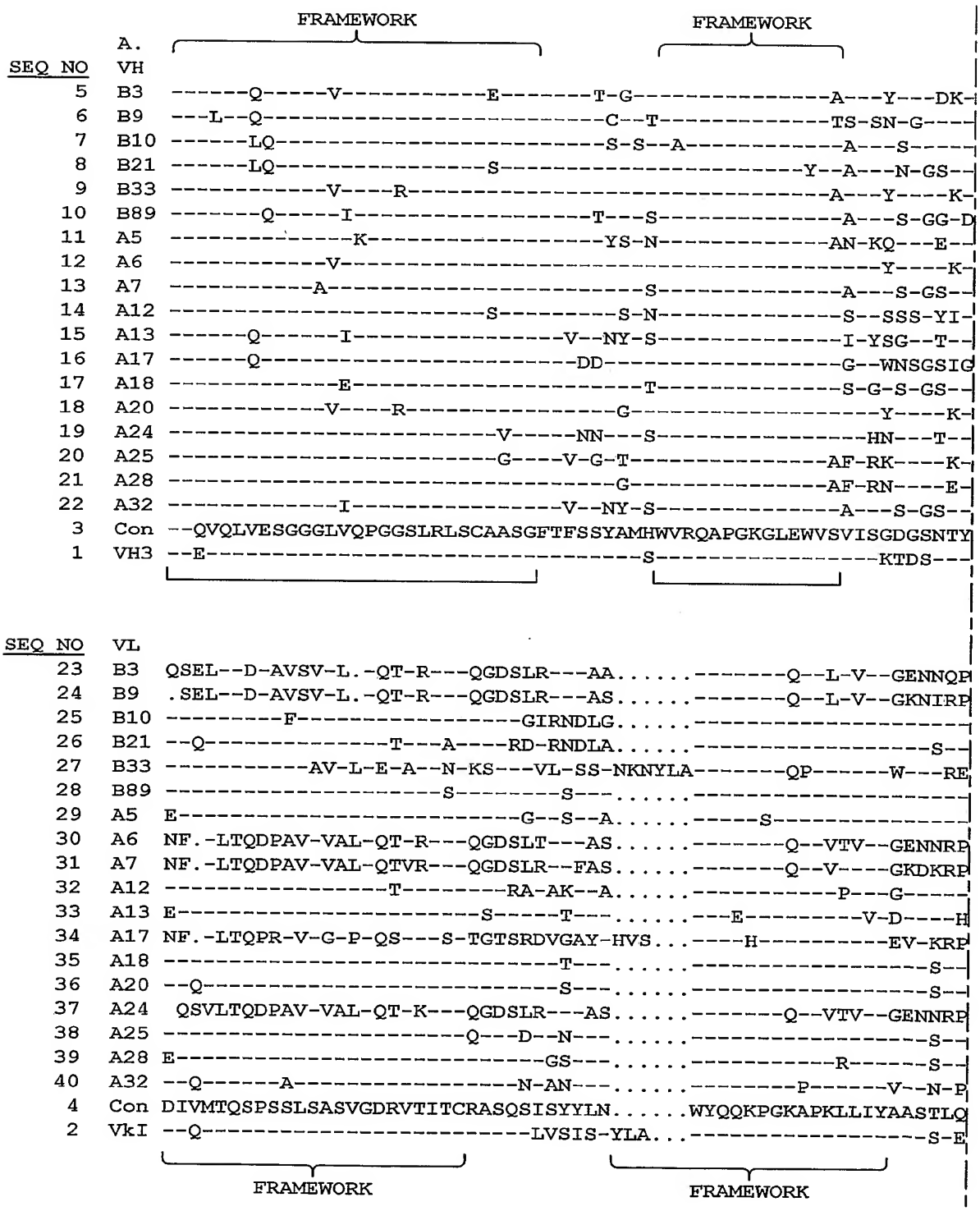

FRAMEWORK
FRAMEWORK
A.
SEQ NO VH
5 B3 ------Q-----V----------E-------T-G---------------A---Y---DK-
6 B9 ---L--Q-------------------------C--T-------------TS-SN-G----
7 B10 ------LQ-----------------------S-S--A------------A---S-----
8 B21 ------LQ----------------S-----------------------Y--A---N-GS--
9 B33 ------------V----R-------------------------------A---Y----K-
10 B89 -------Q-----I-----------------T---S-------------A---S-GG-D
11 A5 ----------------K-----------------YS-N-------------AN-KQ---E--
12 A6 ------------V---Y----K-
13 A7 -----------A-----------------------S-------------A---S-GS--
14 A12 ----------------------S---------S-N-------------S--SSS-YI-
15 A13 ------Q------I----------------V--NY-S-------------I-YSG--T--
16 A17 ------Q-----------------------DD----------------G--WNSGSIG
17 A18 ------------E---------------------T-------------S-G-S-GS--
18 A20 ------------V----R----------------G-------------------Y----K-
19 A24 -------------------------V-----NN---S----------------HN---T--
20 A25 -------------------------G----V-G-T--------------AF-RK----K-
21 A28 ----------------------------------G--------------AF-RN----E-
22 A32 -------------I----------------V--NY-S-------------A---S-GS--
3 Con --QVQLVESGGGLVQPGGSLRLSCAASGFTFSSYAMHWVRQAPGKGLEWVSVISGDGSNTY
1 VH3 --E-------------------------------S----------------KTDS---
SEQ NO VL
23 B3 QSEL--D-AVSV-L.-QT-R---QGDSLR---AA.....-------Q--L-V--GENNQP
24 B9 .SEL--D-AVSV-L.-QT-R---QGDSLR---AS.....-------Q--L-V--GKNIRP
25 B10 ---------F----------------GIRNDLG.....-------------------
26 B21 --Q-------------T---A----RD-RNDLA.....----------------S--
27 B33 -----------AV-L-E-A--N-KS---VL-SS-NKNYLA-------QP------W---RE
28 B89 ---------------------S--------S---.....-------------------
29 A5 E-------------------------G--S--A.....-----S---------------
30 A6 NF.-LTQDPAV-VAL-QT-R---QGDSLT---AS.....-------Q--VTV--GENNRP
31 A7 NF.-LTQDPAV-VAL-QTVR---QGDSLR--FAS.....-------Q--V----GKDKRP
32 A12 ----------------T--------RA-AK--A.....----------P---G-----
33 A13 E----------------------S-----T---.....---E---------V-D----H
34 A17 NF.-LTQPR-V-G-P-QS---S-TGTSRDVGAY-HVS...-----H----------EV-KRP
35 A18 ----------------------------T---.....----------------S--
36 A20 --Q--------------------------S---.....----------------S--
37 A24 QSVLTQDPAV-VAL-QT-K---QGDSLR---AS.....-------Q--VTV--GENNRP
38 A25 ---------------------Q---D--N---.....----------------S--
39 A28 E----------------------GS---.....----------R------S--
40 A32 --Q------A----------N-AN---.....---------P------V--N-P
4 Con DIVMTQSPSSLSASVGDRVTITCRASQSISYYLN.....WYQQKPGKAPKLLIYAASTLQ
2 VkI --Q----------------------LVSIS-YLA...----------S-E
FRAMEWORK
FRAMEWORK

FIG. 5A CONT'D

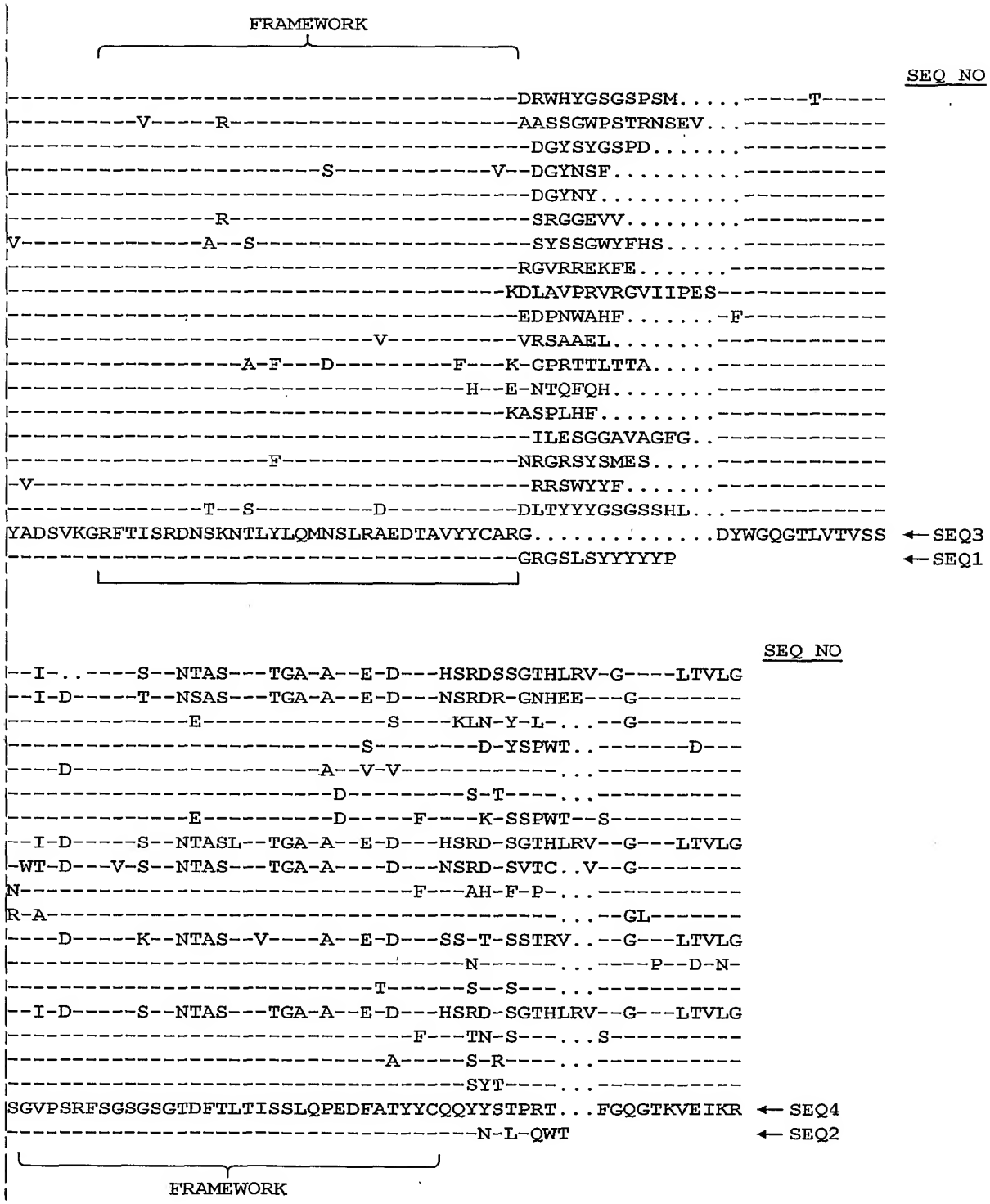


FIG. 5B

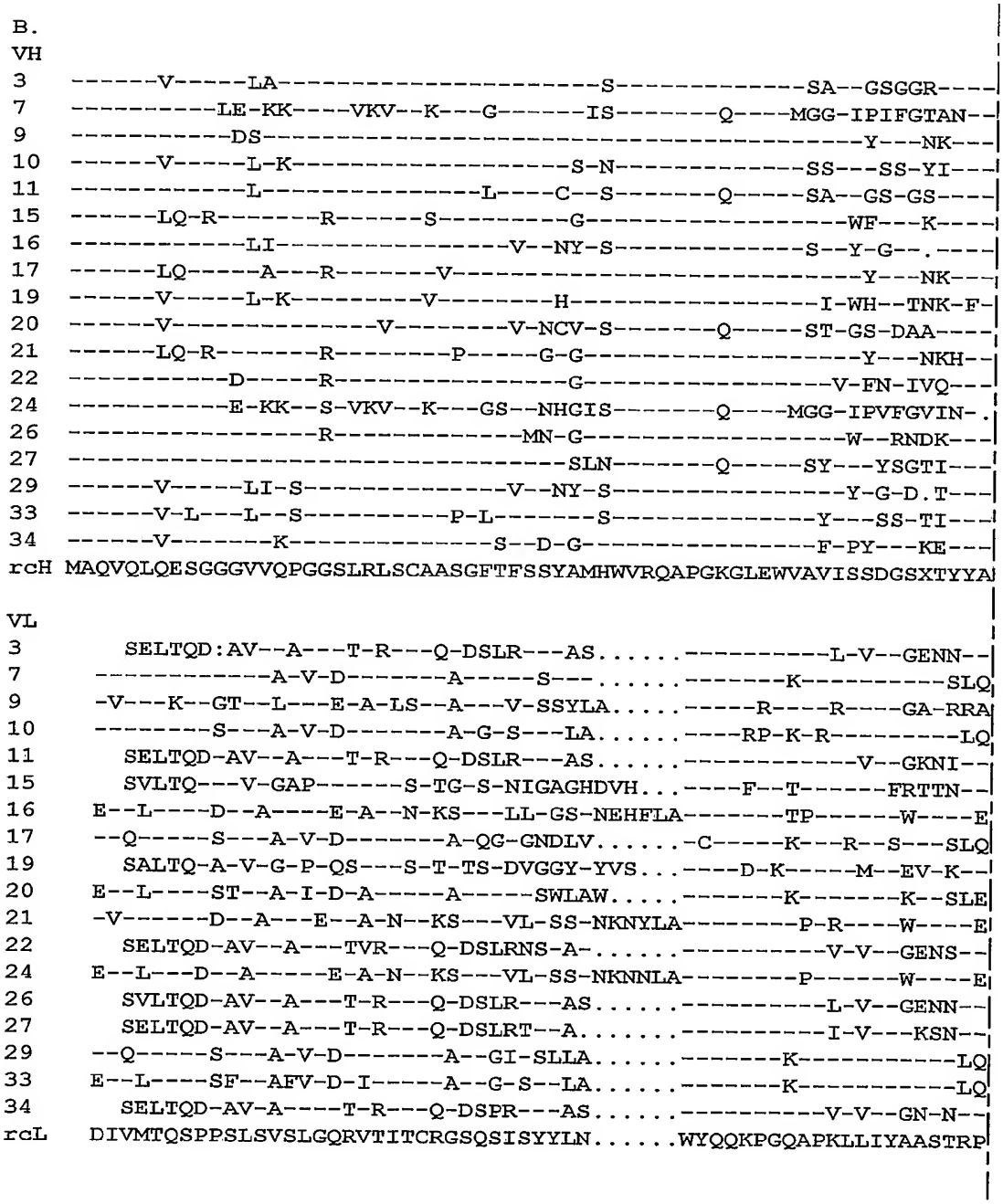
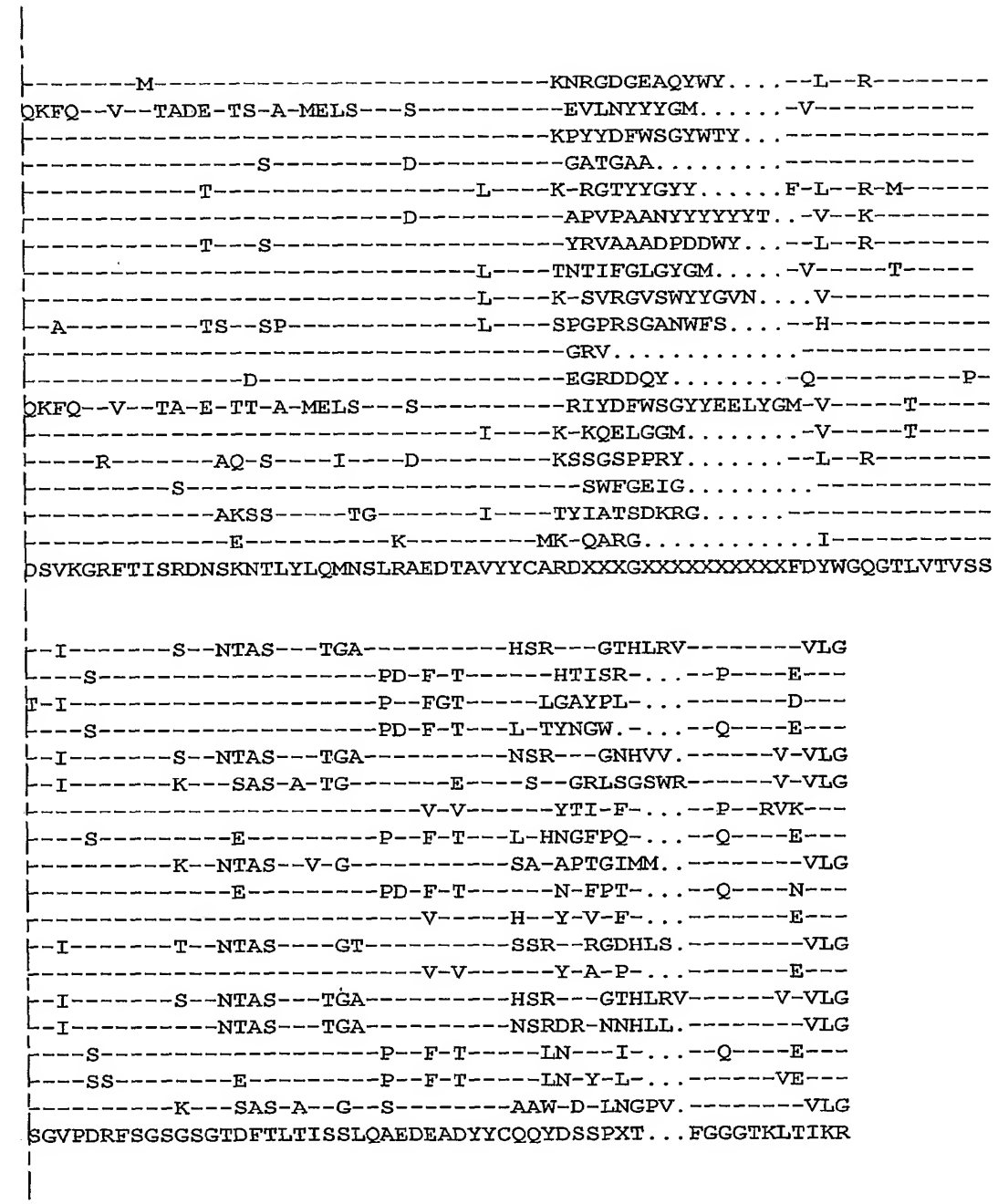

B.
VH
3 ------V-----LA----------S----------SA--GSGGR----
7 ----------LE-KK----VKV--K---G------IS-------Q----MGG-IPIFGTAN--
9 ----------DS----------Y---NK---
10 ------V-----L-K----------S-N----------SS---SS-YI---
11 ----------L----------L----C--S-------Q-----SA--GS-GS----
15 ------LQ-R-------R------S---------G----------WF---K----
16 ----------LI----------V--NY-S----------S--Y-G--.----
17 ------LQ-----A---R-------V----------Y---NK---
19 ------V-----L-K---------V--------H----------I-WH--TNK-F-
20 ------V----------V--------V-NCV-S-------Q-----ST-GS-DAA----
21 ------LQ-R-------R--------P-----G-G----------Y---NKH--
22 ----------D-----R----------G----------V-FN-IVQ---
24 ----------E-KK--S-VKV--K---GS--NHGIS-------Q----MGG-IPVFGVIN-.
26 ----------R----------MN-G----------W--RNDK---
27 ----------SLN-------Q-----SY---YSGTI---
29 ------V-----LI-S----------V--NY-S----------Y-G-D.T---
33 ------V-L---L--S----------P-L-------S----------Y---SS-TI---
34 ------V-------K----------S--D-G----------F-PY---KE---
rcH MAQVQLQESGGGVVQPGGSLRLSCAASGFTFSSYAMHWVRQAPGKGLEWVAVISSDGSXTYYA
VL
3 SELTQD:AV--A---T-R---Q-DSLR---AS.....----------L-V--GENN--
7 ----------A-V-D-------A-----S---.....-------K----------SLQ
9 -V---K--GT--L---E-A-LS--A---V-SSYLA.....-----R----R----GA-RRA
10 --------S---A-V-D-------A-G-S---LA.....----RP-K-R---------LQ
11 SELTQD-AV--A---T-R---Q-DSLR---AS.....----------V--GKNI--
15 SVLTQ---V-GAP------S-TG-S-NIGAGHDVH...----F--T------FRTTN--
16 E--L----D--A----E-A--N-KS---LL-GS-NEHFLA-------TP------W----E
17 --Q-----S---A-V-D-------A-QG-GNDLV.....-C-----K---R--S---SLQ
19 SALTQ-A-V-G-P-QS---S-T-TS-DVGGY-YVS...----D-K-----M--EV-K--
20 E--L----ST--A-I-D-A-----A-----SWLAW.....-------K-------K--SLE
21 -V------D--A---E--A-N--KS---VL-SS-NKNYLA--------P-R----W----E
22 SELTQD-AV--A---TVR---Q-DSLRNS-A-.....----------V-V--GENS--
24 E--L---D--A-----E-A-N--KS---VL-SS-NKNNLA--------P------W----E
26 SVLTQD-AV--A---T-R---Q-DSLR---AS.....----------L-V--GENN--
27 SELTQD-AV--A---T-R---Q-DSLRT--A.....----------I-V---KSN--
29 --Q-----S---A-V-D-------A--GI-SLLA.....-------K----------LQ
33 E--L----SF--AFV-D-I-----A--G-S--LA.....-------K----------LQ
34 SELTQD-AV-A----T-R---Q-DSPR---AS.....----------V-V--GN-N--
rcL DIVMTQSPPSLSVSLGQRVTITCRGSQSISYYLN.....WYQQKPGQAPKLLIYAASTRP

FIG. 5B CONT'D


QKFQ--V--TADE-TS-A-MELS---S----------EVLNYYYGM.....-V----------
QKFQ--V--TA-E-TT-A-MELS---S----------RIYDFWSGYYEELYGM-V-----T-----
DSVKGRFTISRDNSKNTLYLQMNSLRAEDTAVYYCARDXXXGXXXXXXXXXXFDYWGQGTLVTVSS
SGVPDRFSGSGSGTDFTLTISSLQAEDEADYYCQQYDSSPXT...FGGGTKLTIKR

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 June 2004 (03.06.2004)

PCT

(10) International Publication Number
WO 2004/046189 A3

(51) International Patent Classification[7]: C12N 15/13, C07K 16/18, 16/28, C12N 15/10

(21) International Application Number: PCT/GB2003/004964

(22) International Filing Date: 14 November 2003 (14.11.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0226731.8 15 November 2002 (15.11.2002) GB

(71) Applicant *(for all designated States except US)*: **MEDICAL RESEARCH COUNCIL** [GB/GB]; 20 Park Crescent, London W1B 1AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **RABBITTS, Terence, Howard** [GB/GB]; MRC Laboratory of Molecular Biology, Division Of Protein and Nucleic Acid Chemistry, Hills Road, Cambridge CB2 2QH (GB). **CHUNG, Grace** [CN/CN]; Department of Chemical Pathology, The Chinese University of Hong Kong, The Prince of Wales Hospital, Shatin, New Territories, Hong Kong (CN). **TANAKA, Tomoyuki** [JP/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB). **LOBATO-CABALLERO, Maria, Natividad** [ES/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB). **FORSTER, A.** [GB/GB]; MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH (GB).

(74) Agents: **MASCHIO, Antonio** et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report: 16 September 2004

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD FOR GENERATING IMMUNOGLOBULIN GENES

(57) Abstract: The invention provides a method for preparing a double stranded nucleic acid which encodes an immunoglobulin, comprising the steps of: (a) providing a set of three or more overlapping oligonucleotides which anneal to form the + and - strands of a nucleic acid which encodes at least part of an immunoglobulin variable domain; (b) annealing the oligonucleotides; (c) replicating the + and - strands of the nucleic acid formed from the annealed oligonucleotides; and (d) inserting the nucleic acid into an expression vector.

WO 2004/046189 A3

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 03/04964

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12N15/13 C07K16/18 C07K16/28 C12N15/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHENG JULONG ET AL: "Construction and expression of a reshaped VH domain against human CD28 molecules." PREPARATIVE BIOCHEMISTRY & BIOTECHNOLOGY. AUG 2002, vol. 32, no. 3, August 2002 (2002-08), pages 239-251, XP008032678 ISSN: 1082-6068 the whole document	1-22
A	WO 00/54057 A (MEDICAL RESEARCH COUNCIL) 14 September 2000 (2000-09-14) cited in the application the whole document	1-22

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 July 2004	21/07/2004

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Moreau, J

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 03/04964

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	TSE ERIC ET AL: "Intracellular antibody capture technology: application to selection of intracellular antibodies recognising the BCR-ABL oncogenic protein." JOURNAL OF MOLECULAR BIOLOGY. 15 MAR 2002, vol. 317, no. 1, 15 March 2002 (2002-03-15), pages 85-94, XP002247076 ISSN: 0022-2836 cited in the application the whole document	1-22
A	US 6 180 370 B1 (QUEEN C.L.ET AL) 30 January 2001 (2001-01-30) the whole document	1-22

Form PCT/ISA/210 (continuation of second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 03/04964

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 0054057 A	14-09-2000	GB	2344886 A	21-06-2000
		AU	3174700 A	28-09-2000
		AU	4237099 A	14-09-2000
		CA	2277561 A1	10-09-2000
		EP	1166121 A1	02-01-2002
		WO	0054057 A1	14-09-2000
US 6180370 B1	30-01-2001	US	5530101 A	25-06-1996
		US	2003229208 A1	11-12-2003
		US	2004049014 A1	11-03-2004
		US	2004058414 A1	25-03-2004
		AT	251639 T	15-10-2003
		AU	7548196 A	20-02-1997
		AU	671949 B2	19-09-1996
		AU	9172691 A	22-07-1992
		CA	2098404 A1	20-06-1992
		DE	69133326 D1	13-11-2003
		EP	1386932 A1	04-02-2004
		EP	0566647 A1	27-10-1993
		JP	6503963 T	12-05-1994
		JP	3276369 B2	22-04-2002
		KR	231090 B1	15-11-1999
		WO	9211018 A1	09-07-1992
		US	5693761 A	02-12-1997
		US	5585089 A	17-12-1996
		US	5693762 A	02-12-1997
		AT	133452 T	15-02-1996
		AT	183753 T	15-09-1999
		AU	647383 B2	24-03-1994
		AU	5153290 A	13-08-1990
		BG	61095 B2	31-10-1996
		CA	2006865 A1	28-06-1990
		CA	2328851 A1	28-06-1990
		CN	1043875 A ,B	18-07-1990
		CZ	9104186 A3	13-10-1993
		DE	10075005 I1	23-11-2000
		DE	68925536 D1	07-03-1996
		DE	68925536 T2	20-06-1996
		DE	68929061 D1	30-09-1999
		DE	68929061 T2	27-04-2000
		DK	94198 A	16-07-1998
		DK	119191 A	19-06-1991
		EP	0451216 A1	16-10-1991
		EP	0682040 A1	15-11-1995
		EP	0939127 A2	01-09-1999
		ES	2081974 T3	16-03-1996
		ES	2136760 T3	01-12-1999
		FI	108797 B1	28-03-2002
		HK	1014718 A1	14-07-2000
		HR	920500 A1	30-04-1996
		HU	9500236 A3	28-09-1995
		IE	20000331 A1	21-02-2001
		JP	11004694 A	12-01-1999
		JP	2828340 B2	25-11-1998
		JP	4502408 T	07-05-1992